ENCYCLOPEDIA OF

Biodiversity

Fo – Man

VOLUME 3

ACADEMIC PRESS

ENCYCLOPEDIA OF
Biodiversity

VOLUME 3

Fo–Man

EDITORIAL BOARD

Editor-in-Chief

Simon Asher Levin

Associate Editors

Robert Colwell
University of Connecticut
Storrs, Connecticut, USA

Gretchen Daily
Stanford University
Stanford, California, USA

Jane Lubchenco
Oregon State University
Corvallis, Oregon, USA

Harold A. Mooney
Stanford University
Stanford, California, USA

Ernst-Detlef Schulze
Universität Bayreuth
Bayreuth, Germany

G. David Tilman
University of Minnesota
St. Paul, Minnesota, USA

International Editorial Advisors

Dan Cohen
Hebrew University of Jerusalem
Jerusalem, Israel

Rita R. Colwell
National Science Foundation
Arlington, Virginia, USA

Francesco di Castri
National Research Center of France
Montpellier, France

Paul R. Ehrlich
Stanford University
Stanford, California, USA

Thomas Eisner
Cornell University
Ithaca, New York, USA

Niles Eldredge
American Museum of Natural History
New York, New York, USA

Paul Falkowski
Rutgers University
New Brunswick, New Jersey, USA

Tom Fenchel
University of Copenhagen
Helsingoer, Denmark

Diana H. Wall
Colorado State University
Fort Collins, Colorado, USA

Madhav Gadgil
Indian Institute of Science
Bangalore, India

Stephen Jay Gould
Harvard University
Cambridge, Massachusetts, USA

Francesca Grifo
American Museum of Natural History
New York, New York, USA

Masahiko Higashi
Kyoto University (deceased)
Kyoto, Japan

Yoh Iwasa
Kyushu University
Fukuoka, Japan

John H. Lawton
Imperial College at Silwood Park
Ascot, Berks, United Kingdom

Sir Robert May
University of Oxford
Oxford, United Kingdom

Ortwin Meyer
Universität Bayreuth
Bayreuth, Germany

Norman Myers
Consultant in Environment and
Development
Headington, Oxford, United Kingdom

Michael J. Novacek
American Museum of Natural History
New York, New York, USA

Sir Ghillean Prance
Royal Botanic Gardens
Richmond, Surrey, United Kingdom

Michael Rosenzweig
University of Arizona
Tucson, Arizona, USA

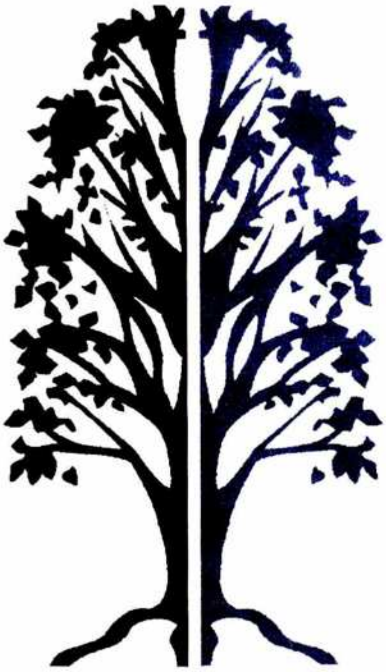
Nigel Stork
Research Center for Tropical Rainforest
Ecology and Management
Cairns, Queensland, Australia

Monica G. Turner
University of Wisconsin
Madison, Wisconsin, USA

Marvalee H. Wake
University of California, Berkeley
Berkeley, California, USA

Brian H. Walker
Commonwealth Scientific and Industrial
Research Organization
Lyncham, Australia

Edward O. Wilson
Museum of Comparative Zoology
Harvard University
Cambridge, Massachusetts, USA



ENCYCLOPEDIA OF
Biodiversity

VOLUME 3
Fo–Man

Editor-in-Chief

Simon Asher Levin

Moffett Professor of Biology
Princeton University
Princeton, New Jersey, USA

ACADEMIC PRESS
A Harcourt Science and Technology Company
SAN DIEGO SAN FRANCISCO BOSTON NEW YORK LONDON SYDNEY TOKYO

This book is printed on acid-free paper.

Academic Press
An imprint of Elsevier Science
525 B Street, Suite 1900, San Diego, California 92101-4495, USA
http://www.academicpress.com

Academic Press
84 Theobalds Road, London WC1X 8RR, UK
http://www.academicpress.com

Library of Congress Catalog Card Number: 02-105903

International Standard Book Number: 0-12-226865-2 set
International Standard Book Number: 0-12-226866-0 Volume 1
International Standard Book Number: 0-12-226867-9 Volume 2
International Standard Book Number: 0-12-226868-7 Volume 3
International Standard Book Number: 0-12-226869-5 Volume 4
International Standard Book Number: 0-12-226864-4 Volume 5

PRINTED IN THE UNITED STATES OF AMERICA
02 03 04 05 06 07 MM 9 8 7 6 5 4 3 2

Dedicated to the memory of three encyclopedia authors,
Takuya Abe, Masahiko Higashi, and Gary Polis, and
their colleagues Shigeru Nakano and Michael Rose,
who perished March 27, 2000 in a tragic boating
accident while on a research trip in Baja California.
Masahiko Higashi was also a member of the
Board of International Editorial Advisors.

Contents

Contents of Other Volumes

CONTENTS OF VOLUME I

CONTENTS OF VOLUME 2

CONTENTS OF VOLUME 4

CONTENTS OF VOLUME 5

Contents by Subject Area

AGRICULTURE

CONSERVATION AND RESTORATION

ECONOMICS OF BIODIVERSITY

ENVIRONMENTAL CONDITIONS AND EFFECTS

EVOLUTION

EXTINCTIONS

GENETICS

GEOGRAPHIC AND GLOBAL ISSUES

HABITATS AND ECOSYSTEMS

HUMAN EFFECTS AND INTERVENTIONS

INVERTEBRATES

MICROBIAL BIODIVERSITY

PLANT BIODIVERSITY

POPULATION ISSUES

PUBLIC POLICIES AND ATTITUDES

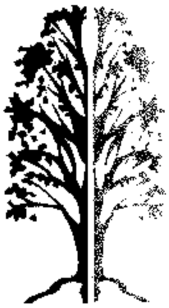
SPECIES INTERACTIONS AND INTERRELATIONSHIPS

SYSTEMATICS AND SPECIES CONCEPT

TECHNIQUES AND MEASUREMENTS

THEORIES AND CONCEPTS OF BIODIVERSITY

VERTEBRATES



Foreword

Biodiversity is the totality of the inherited variation of all forms of life across all levels of variation, from ecosystem to species to gene. Soon after the term was introduced at the first National Forum on Biodiversity in 1986, and after it began its rapid spread around the world, there occurred a reconfiguration in the way much of the science of biology is conceived. Where previously comparative biology had been almost entirely focused on the fundamentals of classification, evolution, behavior, and ecology, now it was augmented by a wide range of analyses from the social sciences. Where taxonomy and biogeography had been marginalized through the middle half of the twentieth century, now they moved back toward center stage. And where extinction had been little more than a phenomenon recognized and lamented, now it became a major concern of science. Much of ecology shifted toward the study of biodiversity's role in the assembly and maintenance of ecosystems. A growing number of economists, political scientists, and bioethicists took up the issue as part of their scholarly agenda. From this mix the discipline of conservation biology was born, and the Society of Conservation Biology became one of the fastest growing organizations in modern science. In 1992 the Rio Summit catapulted biodiversity to global prominence, from which most of the nations of the world endorsed the Convention on Biodiversity and have since used it as a guideline for conservation programs.

The new biodiversity initiative gave organismic and evolutionary biologists a global mission worthy of their science. It confirmed for those who labored in the vineyards that, as medicine is to molecular and cellular biology, the environment is to organismic and evolutionary biology. The first is responsible for personal health, and the second for planetary health. The additional evidence adduced moreover made clear that we are in the midst of an episode of massive extinction, unprecedented since that closing of the Mesozoic Era—and that scientists must lead the attempt to save the Creation.

The articles in the *Encyclopedia of Biodiversity* are unusually eclectic, yet organized by a set of easily articulated goals. They are the following: to carry the systematics and biogeography of the world fauna and flora toward completion; map the hot spots where conservation will save the most biodiversity; orient studies of natural history to understand and save threatened species; advance ecosystems studies and biogeography to create the needed principles of community assembly and maintenance; acquire the knowledge of resource use, economics, and polity to advance conservation programs based on sustainability; and enrich the ethic of global conservation in terms persuasive to all.

The road ahead, down which we must urgently travel, will be smoothed by the exponential growth of information and a growing public awareness and support. The *Encyclopedia of Biodiversity* will serve as an important knowledge base to guide this supremely important effort.

Edward O. Wilson
Museum of Comparative Zoology
Harvard University

Preface

The science of biodiversity has become the science of our future. Our awareness of the disappearance of biodiversity has brought with it a long-overdue appreciation of the magnitude of our loss, and a determination to develop the tools to protect our future. This encyclopedia brings together, for the first time in its completeness, study of the dimensions of diversity with examination of the services that biodiversity provides, and measures to protect it.

The entries in the encyclopedia have been arranged alphabetically but the coverage is designed functionally. At the core is a comprehensive survey of biodiversity, across taxonomic groups and ecological regions. The emergence of biodiversity is then placed in an evolutionary perspective, as background for an understanding of current trends. Particular attention is given to the loss of services—for example, in fisheries, forestry and climate mediation—that are derived from natural systems. These are placed in an economic framework through a comprehensive set of papers that address problems of valuation, costs, and benefits, and develop a framework for prioritizing actions. Finally, a review is given of institutions and other mechanisms that exist and are needed for the preservation of biodiversity and, with it, the services that humans derive from nature.

The background for understanding biodiversity is to be found in the fossil record, and in the evolutionary patterns and trends that it reveals. The encyclopedia hence discusses these patterns, the origins of biodiversity, the effects of geological events, the mechanisms of evolution, and the uniqueness of the evolutionary process, with implications for conservation and restoration. The essential processes in macroevolution are those of speciation and extinction, which together govern the dynamics of diversity at higher levels of organization. These are given extensive coverage, both from a mechanistic and from a historical perspective, and provide an essential context for understanding the rest of the contributions.

The classification of organisms into species and higher taxa, and the elucidation of the mechanisms of natural selection, were the essential intellectual advances that allowed the development of the science of biodiversity. Carl Linnaeus introduced a systematic framework for understanding phylogenies, which continues to provide the foundation for evolutionary studies today; and Charles Darwin's great legacy—the theory of evolution by natural selection—is the essential organizing principle for understanding the processes that gave rise to the patterns Linnaeus recognized. The encyclopedia provides unmatched taxonomic coverage of the organization of diversity into taxonomic groups and complements that with an extensive examination of ecosystems by biogeographic region and by functional type. These chapters elucidate latitudinal trends, life zones, species–area relationships and the distribution of diversity within and among ecological communities.

Throughout the core chapters, there is a healthy balance between empirical facts and conceptual theories. Such theories help to illuminate principles that cross systems and levels of organization, and transform the study of biodiversity into a science. Basic ecological constructs, such as the habitat and the niche, are given extensive treatment, as are key ecological mechanisms such as competition, predation, herbivory, parasitism and mutualism. These treatments are complemented by exploration of fundamental evolutionary mechanisms related to local differentiation, aspect diversity, sex, and recombination, and especially theories of extinction.

With these foundational chapters in hand, one can turn to the contemporary problems in biodiversity and compare today's rapid rates of change to the historical patterns. Key chapters examine agriculture, fisheries, and forests, their importance to human needs, and their status and trends in response to changing land-use patterns, population growth, overexploitation, and climate change. Threatened and endangered species are discussed in detail, with relation to the consequences of the spread of invading species.

The utilization of nature's bounty for food, fiber and fuel provides some of the most obvious benefits of biodiversity to humanity. Equally important, however, are the things that are less well appreciated: the potential for the discovery of new pharmaceuticals that can improve human health, the role of biodiversity in pollinating crops and wild species, and the importance of natural systems in regulating climate, mediating nutrient fluxes, and sequestering carbon as well as toxic materials. Each of these services provides humans with direct and indirect benefits, and somehow we must find ways to weigh these benefits, along with the ethical and aesthetic values we place on natural systems and biodiversity, to provide priorities for action. Only recently have economists recognized the importance of such issues as intellectual challenges essential to our survival on the planet. Much of biodiversity is exploited by humanity as part of a global commons, in which one does not pay in fair measure for extracting parts, or affecting the commons otherwise through land use or pollution. Economists have come to realize, along with ecologists and others, the magnitude of the externalities involved. When such externalities are involved, the market does not function as it must to maintain the resource, and new measures are needed if the sustainability and resilience of these resources are to be preserved. There is as yet, however, no ecological equivalent to the power of financial institutions, such as the Federal Reserve Board in the United States, to modify individual incentives sufficiently to maintain regional or global stability in the system of interest. In this encyclopedia, some of the most enlightened and thoughtful economists turn their attention to the economic challenges, and discuss the mechanisms and institutions that might be needed.

Together, the state-of-the-art entries in this encyclopedia tell an exciting story of how biodiversity arose, continues to arise, and is maintained. It is a story of a complex, self-organizing system—the biosphere—whose pieces can be examined individually, but cannot be understood outside the context of the whole. It is also a story of the coevolution of the biosphere and *Homo sapiens*, the first species whose own activities can feed back to influence the evolution of the biosphere on time scales that could lead to its own demise. The articles in the encyclopedia can be used as material for a wide spectrum of courses, tracing the history of the emergence of biodiversity from its origins to the challenges we face today.

This has been a massive effort, but one of the most rewarding I have ever undertaken. So many people have played a role that it is difficult to know where to begin. The project began through the initiative of Scott Bentley at Academic Press, and then was managed flawlessly by Chris Morris at AP. I cannot recall ever having dealt with an editor who operated more professionally than Chris, who combined a true vision and enthusiasm for the project with a sense of economic realities and the energy and insight to make the whole project work. In this he was ably assisted by outstanding Academic Press staff, especially Naomi Henning, Nick Panissidi, and Ann Marie Martin. At every step, it has been a pleasure to work with Academic Press, and I especially single out Chris for his fantastic and scholarly efforts.

At the next stage, the Editorial Boards were terrific in generating and commenting on ideas, suggesting authors, and critiquing contributions. More than 400 authors then adopted our view of the importance of the project, accepted the task of writing, and produced timely and comprehensive articles that make this Encyclopedia like no other source available today. To all of these, I extend my thanks and congratulations.

And finally, special acknowledgment and gratitude are due my wife, Carole, and my assistant, Amy Bordvik. Carole put up with the late nights and obsessiveness that were essential to the process, and Amy worked tirelessly and without complaint through the whole long process. To them, any expression of thanks is insufficient.

Simon Levin
Princeton University

References

Simon A. Levin (1999). *Fragile Dominion,* Perseus Books, Reading, Massachusetts

Edward O. Wilson (1992). *The Diversity of Life.* Norton. New York, New York.

Guide to the Encyclopedia

The *Encyclopedia of Biodiversity* is a comprehensive study of the topic of diversity in the natural world, contained within the covers of a single unified work. It consists of five volumes and includes 313 separate full-length articles by leading international authors.

Each article in the encyclopedia provides a comprehensive overview of the selected topic to inform a broad spectrum of readers, from research professionals to students to the interested general public. In order that you, the reader, will derive the greatest possible benefit from the *Encyclopedia of Biodiversity*, we have provided this Guide. It explains how the encyclopedia was developed, how it is organized, and how the information within it can be located.

ENTRY SELECTION

This encyclopedia was conceived with the goal of providing a complete description of all the issues contained within, or impacting upon, the field of biodiversity. To that end, a thorough and systematic method of entry selection was devised for the work.

To begin the selection process, the project's chief editor, Simon Levin, prepared a bibliography of leading source materials in the field, including books, journal articles, conference proceedings, Websites, and so on. Then the reference staff of Academic Press combed through these materials to develop a list of potential article topics for the encyclopedia. This preliminary list was refined and approved by Prof. Levin; at this point the number of possible entries was approximately twice as large as the eventual total in the published encyclopedia.

The entry list was then provided to all the associate editors and the international editorial board for their evaluation. Their mandate was to read through the list and rate each topic on a numerical scale according to how important they deemed it to be for inclusion in the encyclopedia. The editors were also encouraged to recommend new topics not on the existing list, and to make other comments on the list as appropriate. A number of additions to the entry list emerged from this process.

The editors' ratings and comments were returned to Academic Press for scoring, and an overall tabulation was created that indicated the consensus of the group as to the priority of each topic. Then the list was sent to Prof. Levin for a final evaluation in which he made "tie-breaker" decisions for certain topics on which the editors' vote was split, and also other adjustments based on his expert judgment. The result was a working entry list of about 325 topics which, after some attrition and the combining of related topics, resulted in the final table of contents of 313 articles.

ORGANIZATION

The *Encyclopedia of Biodiversity* is organized in a single alphabetical sequence by title. Articles whose titles begin with the letters A to C are in Volume 1, articles with titles from D through Fl are in Volume 2, then Fo through Man in Volume 3, Mar through Q in Volume 4, and R to Z in Volume 5.

Volume 5 also includes a complete subject index for the entire work, an alphabetical list of the authors who contributed to the encyclopedia, and a glossary of key terms used in the articles.

TABLE OF CONTENTS

A complete table of contents for the *Encyclopedia of Biodiversity* appears at the front of each volume. This alphabetical list of article titles (see p. vii) is followed by a second contents list (p. xix) in which the titles are listed according to their subject area within the overall field of biodiversity.

Articles are classified in 20 different subject areas, including not only core disciplines of biodiversity such as evolution, speciation, populations, extinction, and ecosystems, but also areas that link biodiversity to other disciplines, such as environmental science, agriculture, public policy, and economics.

ARTICLE TITLES

Article titles generally begin with the key term describing the topic, and have inverted word order if necessary to begin the title with this term. For example, "Archaea, Origin of" is the article title rather than "Origin of Archaea," "Grazing, Effects of" is the title rather than "Effects of Grazing," and so on with other titles such as "Species, Concepts of," "Mammals, Biodiversity of," "Pollinators, Role of," and so on. This is done so that the reader can more easily locate a desired topic. For example, eight different articles on endangered groups (e.g., endangered birds) appear in succession in the "En-" section of the encyclopedia.

INDEX

The index appears as the last element of Volume 5. Subjects are listed alphabetically and indicate the volume and page number where information on this topic can be found. In addition, the table of contents by subject area also functions as an index, since it lists all the topics covered in a given area; e.g., the encyclopedia has 26 different articles dealing with invertebrates.

ARTICLE FORMAT

Articles in the *Encyclopedia of Biodiversity* are arranged in a standard format, as follows:

- Title and Author
- Outline
- Glossary
- Defining Statement
- Main Body of the Article
- Cross-References
- Bibliography

OUTLINE

Entries in the encyclopedia begin with a topical outline that indicates the general content of the article. This outline serves two functions. First, it provides a preview of the article, so that the reader can get a sense of what is contained there without having to leaf through the pages. Second, it serves to highlight important subtopics that are discussed within the article. For example, the article "Greenhouse Effect" includes subtopics such as "Climatic Consequences: Global Warming" and "Climate Change and Biodiversity."

The outline is intended as an overview and thus it lists only the major headings of the article. In addition, extensive second-level and third-level headings will be found within the article.

GLOSSARY

The Glossary section contains terms that are important to an understanding of the article and that may be unfamiliar to the reader. Each term is defined in the context of the article in which it is used. The same term may appear as a glossary entry in different articles, with the details of the definition varying slightly from one article to another. The encyclopedia includes approximately 2,500 glossary entries. For example, the article "Mangrove Ecosystems" has the following entry:

> **aerenchyma** A spongy plant tissue composed largely of air spaces enabling gas exchange to take place by diffusion in underground mangrove roots.

In addition, Volume 5 has a comprehensive glossary that presents the core vocabulary of biodiversity in one A-Z list. This section can be consulted for definitions of unfamiliar terms not found in the individual glossary for a given article.

DEFINING STATEMENT

The text of each article in the encyclopedia begins with a single introductory paragraph that defines the topic under discussion and summarizes the content of the article. For example, the article "Agriculture, Sustainable" begins with the following statement:

> Sustainable agriculture describes a food and fiber production system that is economically viable, environmentally safe, and socially acceptable over long periods.

CROSS REFERENCES

The entry list for *Encyclopedia of Biodiversity* has been constructed so that each entry is supported by one or more other entries that provide additional information. Therefore all articles in the encyclopedia have references to other articles. These cross references appear at the conclusion of the article text. They indicate articles that can be consulted for further information on the same issue, or for pertinent information on a related issue. The encyclopedia includes a total of about 1,750 cross references to other articles. For example, the article "Biodiversity-Rich Countries" contains the following list of references:

> Biodiversity as a Commodity • Deforestation • Economic Growth and the Environment • Indigenous Peoples, Biodiversity and • Social and Cultural Factors • Tropical Ecosystems

BIBLIOGRAPHY

The Bibliography section appears as the last element in an article. Entries in this section include not only print sources but relevant Websites as well.

The bibliography entries in this encyclopedia are for the benefit of the reader and do not represent a complete list of all the materials consulted by the author in preparing the article. Rather, the sources listed are the author's recommendations of the most appropriate materials for further research on the given topic. For example, the article "Fires, Ecological Effects of" lists as references (among others) the works *Fire and Plants, Fire in the Environment, Fire in the Tropical Biota,* and *The Role of Fire in Mediterranean Ecosystems.*

COMPANION WORKS

Encyclopedia of Biodiversity is part of a continuing program of multivolume reference works published by Academic Press. This program encompasses many different areas of science, ranging from organismal biology (e.g., *Encyclopedia of Dinosaurs, Encyclopedia of Microbiology*) to biomedical topics (*Encyclopedia of Reproduction, Encyclopedia of Stress*), to physical science (*Encyclopedia of the Solar System, Encyclopedia of Volcanoes*) to social and political issues (*Encyclopedia of Applied Ethics, Encyclopedia of Creativity, Encyclopedia of Nationalism, Encyclopedia of Violence, Peace, and Conflict*).

For information on these and other Academic Press reference titles, please see the Website at:

www.academicpress.com/reference/

FOOD WEBS

Gary R. Huxel and Gary A. Polis[†]
University of California, Davis

GLOSSARY

community The most practical definition is a set of species that interact at a given location.

connectivity web This type of food web illustrates only feeding links without reference to strength of interaction or energy flow.

detrital shunts Energy and nutrients from the saprovore web reenter the plant herbivore predator food web when detritivores are eaten by predators that also eat plants, herbivores, or other predators.

donor control Consumer population growth is affected by their resources but consumers do not affect the renewal rate of these resources and hence cannot depress their resources.

ecosystem A set of one or more communities and their abiotic environment.

energetic web This type of food web quantifies the amount of energy (or material) that flows across links joining species.

food or biomass pyramid A graphic representation of the energy or biomass relationships of a community, in which the total amount of biomass, or total amount of energy available, at each successive trophic level is proportional to the width of the pyramid at the appropriate height.

food chain A representation of the links between consumers and their resources, for example nutrients → plant → herbivore → carnivore. In these representations, energy or material flows up the chain in a linear fashion. In addition, a food chain can be a linear set of species within a food web.

food web A representation of feeding relationships in a community that includes all the links revealed by dietary analysis.

functional or interaction web This type of food web quantifies the strength of interaction between species linked using data from manipulative experiments.

recipient control Consumers substantially depress populations of their resources.

spatial subsidies Input from other habitats of organic carbon, nutrients, and prey or the movement of consumers. These resources can influence greatly the energy, carbon, and nutrient budget of recipient habitats. In general, nutrient inputs (nitrogen, phosphorus, and trace elements) increase primary productivity; detrital and prey inputs produce numerical responses in their consumers.

trophic level An abstract classification to describe subsets of species that acquire energetic resources in a similar way on a subset of species (e.g., top carnivores feed on primary carnivores which feed on herbivores which feed on primary producers). In natural systems, most species do not feed strictly on the

---

[†] Deceased.

"trophic level" below them, making the trophic level concept a difficult term to assign operationally to species.

KNOWLEDGE OF FOOD WEB structure and dynamics is central to our understanding of almost all aspects of population and community ecology. By their very nature of representing feeding relationships between species, food webs have the capacity to embody the rich complexity of natural systems. In fact, most important interactions (e.g., competition, predation, and mutualism) cannot be isolated from a food web context.

I. INTRODUCTION

Food webs occupy a central position in community ecology. Charles Darwin introduced the concept of an entangled bank in which he envisioned many kinds of species interdependent on each other in a complex manner governed by "laws acting around us." In the simplest context, food webs incorporate the two factors that, a priori, one would consider most fundamental to the success of any one species: resources and enemies. All species must acquire resources (food or nutrients) and suffer energy losses or mortality from predators (Fig. 1). The abundance and success of any species is thus a product of these feeding interactions. This inclusion of such "bottom-up" (productivity and resources) with "top-down" (consumption) factors largely determines the distribution and abundance of almost every species on the planet. In particular, freshwater ecologists have enjoyed notable success by concurrently studying the interaction between these variable factors on the regulation of plant and animal abundance and thus the structure of freshwater communities. This research shows the rich dynamical outcomes that can occur when predation and productivity vary and interact within a food web (Fig. 2).

Many important advances have arisen from analyses that concurrently incorporate more than one interaction in a food web: keystone predation and herbivory, the intermediate predation and disturbance hypotheses, the size-efficiency hypothesis, trophic cascades, intraguild predation, apparent competition, and the recognition of the importance of indirect effects. The outcome of virtually all interactions within a community can be modified, directly and indirectly, by other members of the food web. This insight penetrates to all areas of community ecology. For example, the results of experiments must be interpreted carefully for at least two reasons. First, indirect effects, moderated by other species in the web, may exert large and sometimes contradictory effects to the direct effects of the manipulation. Thus, under some food web configurations, removal of a predator may directly increase the level of its prey or may actually cause the prey to decrease because of indirect interactions. Second, changes in species dynamics putatively caused by one factor may actually be a product of a second process.

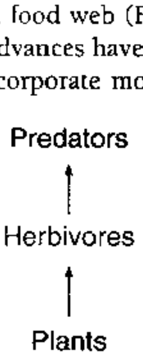


FIGURE 1 Food chain.

II. TYPES OF FOOD WEBS

Food web research has grown at a tremendous rate and taken a diversity of forms. Not surprisingly, ecologists have diverged in their methods, emphases, and approaches. Nevertheless, trophic relationships in communities can be delineated in three basic ways. Paine (1980) and Polis (1991) distinguished three types of food webs that evolved from ecological studies (Fig. 3). The first is the classic food web, a schematic description of connectivity specifying feeding links. Such connectivity webs simply demonstrate feeding relationships. Examples of these are the early food webs of Forbes and Summerhayes and Elton (Fig. 3). The second web type is also descriptive, quantifying the flow of energy and matter through the community. These energetic webs quantify the flow of energy (and/or materials) between trophically connected species. Examples of this type of food web include intertidal communities in Torch Bay, Alaska, and Cape Flattery, Washington (Paine, 1980). The third type use experiments to dissect communities to identify strong links and dynamically important species. Such interaction or functional webs demonstrate the most important connections in an ecosystem (Fig. 3). These food webs depict the importance of species in maintaining the integrity and stability of a community as reflected in its influence on the growth rates of other species. They require experimental manipulations of the community (e.g., by removal or addition of particular species). In the following sections, we discuss the strengths and weaknesses of each ap-

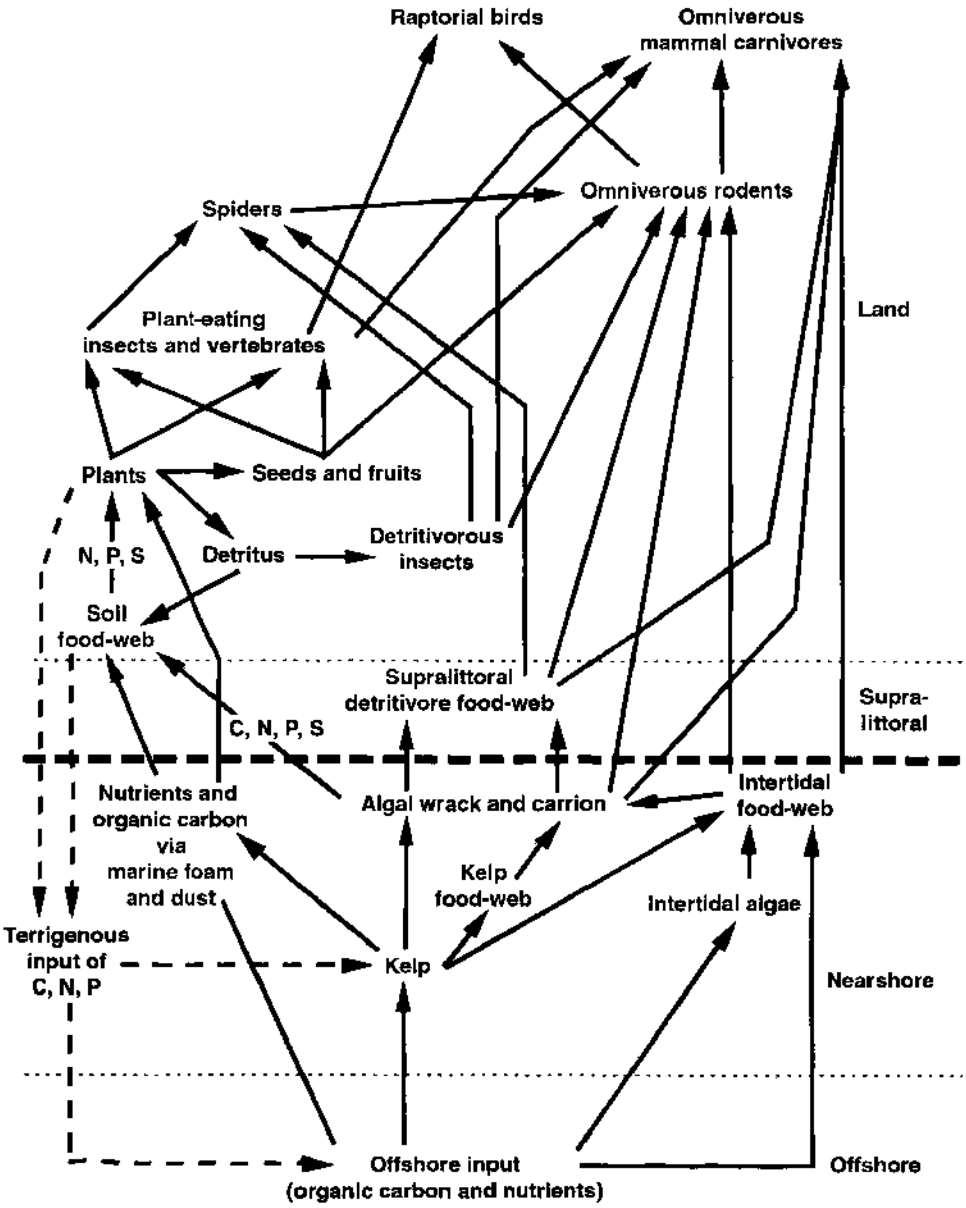


FIGURE 2 Food web.

proach. Of the three, only the last two have contributed substantially to our understanding of natural systems.

A. Connectivity Webs

Connectivity webs are representations of "who eats whom" without inference to the strength or type of interaction and energy flow (Fig. 3). Early food webs were constructed for essentially two reasons: (i) to depict the interconnectivity of natural systems and (ii) to examine issues of "the balance of nature," i.e., to analyze how harmony is maintained through complex predatory and competitive interactions within communities (Forbes, 1887). Such an approach was applied to agricultural systems to examine pests and possible food web manipulations to control pests. As early as the 1880s, beetles were introduced into the United States to control agricultural pests. Such control then benefited crop plants via an indirect interaction (predator pest prey crop) (following the success of Vedalia, a coccinellid beetle, in controlling cottony-cushion scale

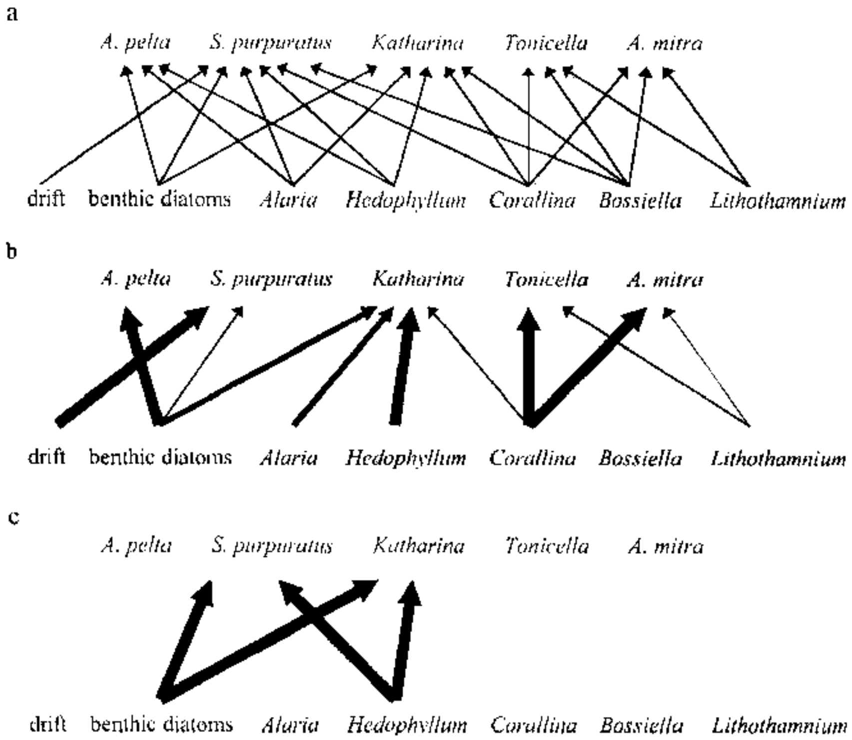


FIGURE 3 Three conceptually and historically different approaches to depicting trophic relationships, illustrated for the same set of species. The connectedness web (a) is based on observation, the energy flow web (b) on some measurement and literature values, and the functional web (c) on controlled manipulation. Used with permission of Blackwell Scientific Publications.

in California in 1888, about 50 more coccinellids were introduced in the 1890s).

The knowledge required to construct connectivity webs is straightforward: An approximate, qualitative knowledge of who eats whom is all that is necessary to produce a simple food web, whereas experimental manipulations or quantitative measurements are necessary to construct webs of interaction or energy flow. Consequently, connectivity webs most frequently represent trophic interactions in communities and have received the most attention. Hundreds of such webs slowly accumulated over a century. They were useful to illustrate, in a totally nonquantitiative manner, the feeding interactions within a specific community. Different scientists constructed webs of different diversity, complexity, and resolution, depending on their knowledge of the system and bias or understanding of particular groups. For example, some may emphasize birds and lump all insects as one group. Others will divide the insects into scores of groups and represent one or two bird species.

In the 1970s and 1980s, many theoretical and statistical studies were performed on connectivity webs cataloged from the literature to determine similarities and natural patterns among them. Empirical generalizations were abstracted from data of published connectivity webs. These "natural patterns" largely agreed with predictions made by early food web models. These models showed that food webs were constrained to be quite simple: Each species ate few species and had few predators; the total length of the number of links in a typical food chain was short, usually two or three; omnivory was very rare; and there were a few other patterns. Early modelers argued that the congruence of patterns from the cataloged webs validated the predictions of their models. They thus claimed that their Lotka–Volterra models were heuristic and represented processes that structure real communities. For example, the addition of omnivory to model food webs causes webs to be unstable dynamically and exhibit relative low persistence (time before species are lost). Thus, these models make the prediction that omnivory should be relatively rare in those webs that persist in nature. Comparison of omnivory in cataloged webs relative to its frequency based on chance shows that omnivory is statistically rare in real webs, as predicted by models. The same

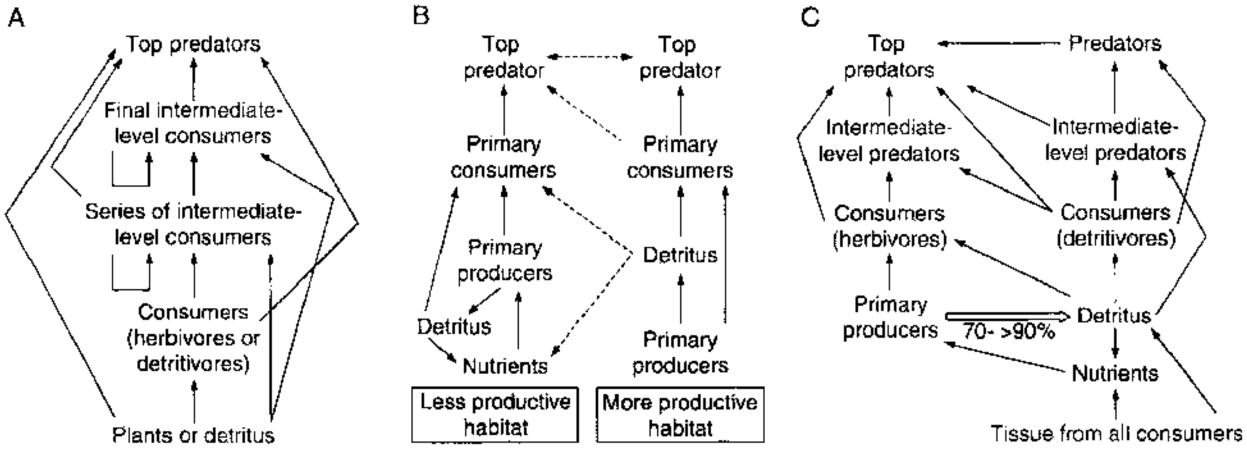


FIGURE 4 Food web showing aggregation within some trophic levels but not others. (A) The dynamics of omnivory; (B) spatial subsidy; (C) detrital shunts.

general approach was used to validate other predictions of model webs, e.g., short chain lengths.

Thus, modelers soon "explained" these empirically derived patterns. Although these studies, and the connectivity approach, make good food web diagrams, they are flawed to such a great a degree that today such analyses are viewed as providing little understanding of natural communities. There are many reasons why this is so, of which only a few are mentioned here:

1. Most vastly under-represent the species diversity in natural communities. Most communities have hundreds to thousands of species, but these webs would represent <10–30 species on the average. As a consequence, most connectivity webs have severe problems with "lumping" species and taxonomic biases. Some trophic levels are distinguished by species (e.g., birds or fish), whereas other groups suffer a high degree of aggregation, e.g., all species of insect or annual plants are represented as one super-species—"insects" or "plants" (Fig. 4).

2. Most species are highly omnivorous, feeding on many resources and prey that each have a distinct trophic history and are often at different trophic levels. Because diet is very difficult to delineate, most connectivity webs greatly underrepresent the true nature of omnivory. This poses several fundamental problems.

3. Connectivity webs typically only offer a static view of the world and webs are usually idealized representations that show all linkages that occur over large spatial and temporal scales. Therefore, much of the important variability and changes due to local environmental conditions are lost. However, studies that compare changes in connectivity over time and space and across environmental gradients (such as those by Mary Power and her group on the Eel River) can provide important insight into community structure and dynamics. One can view connectivity webs as a first step in examining the interactions in communities (i.e., performing "natural history" studies), to be followed by quantification of the fluxes of energy and nutrients (as in energetic webs).

B. Energetic Webs

Starting with the classic studies of Elton, Summerhayes, and Lindeman, food web studies turned toward quantifying flows of energy and nutrients in ecosystems and the biological processes that regulate these flows. This approach is an alternative to connectivity webs to describe trophic connectedness within communities. This "process-functional" approach explicitly incorporates producers, consumers, detritus, abiotic factors, flow out of a system, and the biogeochemical recycling of nutrients. It views food webs as dynamic systems in time and space. Such an approach necessitated analyzing energy and material fluxes in order to understand the behavior of ecosystems. Thus, a typical analysis would quantify the amount of energy or matter as it travels along different pathways (e.g., plants → consumers → detritus → decomposers → soil). For example, the tracking of energy and DDT through a food web in a Long Island estuary enabled researchers to study bioaccumulation effects on top predators.

The use of energetic webs has provided a rich understanding of the natural world and allowed us to understand much about ecosystems. Several important processes are included in energetic webs. First, they quantify energy and material pathways and key species or processes that facilitate or impede such flows. Second, they include an explicit recognition of the great importance of detritus, a subject virtually ignored in connectivity webs. (10 to >90% of all primary productivity from different habitats immediately becomes "dead" organic detritus rather than being eaten by herbivores). Third, this approach recognized that a great amount of energy, nutrients, and prey originated outside the focal habitat, which is a key insight to understand natural communities. Thus, energetic webs show how ecosystems function and which species dominate biomass and energy.

Beginning with Lindeman, researchers began to examine the efficiency of transfer from prey species to predator species. It was found that energy transfer is generally inefficient with only about 5–15% of the energy of prey species being converted to energy of predators. Peter Yodzis used this information to suggest that the length of food chains within a community would be set by the amount of energy entering into the base of the chain. This argument was in opposition to Pimm and Lawton's suggestion that food chain length is set by the resilience of the chain. By resilience, Pimm and Lawton, using Lotka–Volterra models, meant the estimated time for model food chains to recover from some disturbance. They argued that frequent disturbances (relative to growth rates of species) would result in shorter food chain lengths. Furthermore, early studies examining the influence of primary productivity (thus, the amount of energy entering a food chain) did not support the hypothesis that food chain length was governed by energy transfer efficiency. However, recent reexaminations of Pimm and Lawton's work suggest that two factors influenced their results—density-dependent regulation of the basal trophic level and food chain structure (the lack of omnivory in their models). Moreover, recent studies of the role of energy efficiency have found that decreases in productivity result in shorter maximum food chains. Thus, the relative role of resilience versus energy transfer in regulating the length of food chains is still debated.

One outcome of the argument for the role of energy transfer as the main governing factor of food chain length is a body of work that examines differences in energy efficiency among organisms. For example, carnivores are found to have greater efficiency than herbivores. Additionally, invertebrate ectotherms have greater efficiencies than vertebrate ectotherms, which in turn are more efficient than endotherms. Yodzis and Innes used this information (and relative body sizes) to parameterize nonlinear predator–prey models.

In summary, the analysis of energy and matter flow is necessary and central to understanding the dynamics of populations and communities. The success of a population is always strongly related to the energy and biomass available to it. Consequently, it is difficult or impossible to understand the dynamics and structure of food webs and interacting populations without incorporating energy flow from below. However, this energetic approach per se, although necessary, is not sufficient by itself to understand the dynamics of communities because energy flow and biomass production are functions of interactions among populations within the food web. The transfer of energy and matter becomes complicated as they pass through the many consumers that populate community food webs. For example, increasing the amount of nutrients to plants may increase the biomass of each consumer in the web or may just increase the biomass of a subset of consumers (e.g., only the plants, plants and herbivores, or only the herbivores), depending on the relationship between consumers and their resources. Because of these considerations, pathways must be placed in the context of "functional" food webs to understand the dynamics of energy and material transfer.

C. Functional or Interaction Webs

Functional or interaction webs use experiments to determine the dynamics within a community. Starting with Connell and Paine, empiricists began to use experiments to examine communities and food webs to discover which species or interaction most influenced population and community dynamics. They manipulated species that natural history or energetic analyses suggested were important. They used either "press" (continual) or "pulse" (singular) experiments to manipulate populations of single species and then followed the response of other species within the food web. The philosophy of these studies was to simplify the complexity of natural systems with the assumption that many species and links between species were unimportant to dynamics. Paine tested this assumption and found that indeed many links between species were weak (essentially zero).

Experimental analyses of food webs are designed to identify species and feeding links that most influence population and community dynamics. These alone are placed into an "interaction web" that, in theory, encompasses all the elements that most influence the distribution and abundance of member species. However, unlike connectivity webs, key species are identified through experiments rather than diet frequency or energy transfer. The initial process of choosing certain species and interactions for experiments and excluding others is subjective, optimally based on strong intuition and a rich understanding of natural history. As the researcher learns more, some elements are discarded and others are subject to further experimentation. Eventually, the community is distilled into an interaction web, a subset including only species that dominate biomass and/or regulate the flow of energy and matter.

This approach has been used by experimental and theoretical ecologists to produce a rich understanding of the processes that most influence their communities. They have been remarkably fruitful and have introduced many food web paradigms that go to the center of ecology, e.g., keystones species, the intermediate disturbance or predation hypothesis, the size-efficiency hypothesis, top-down and bottom-up control, trophic cascades, and apparent competition.

However, this approach is not without limitations. Three major problems stand out. First, many statistical shortcomings can beset experimental manipulation of food webs. For example, replications are commonly difficult (time-consuming and expensive) and therefore experiments often lack the statistical power necessary to avoid type II statistical errors (significant biological differences exist among treatments but low sample size precludes their detection statistically). Second, the number of possible experiments is almost infinite. Which ones should be conducted, and which species should be manipulated?

The third and perhaps most troublesome problem is that experiments isolate a subset of species and links from the community food web, largely ignoring how manipulations interact with the remainder of the community. Thus, unobserved indirect or higher order interactions may exert important effects on the dynamics of experimental species and, in theory, make the outcome of experiments indeterminate. For example, predators are thought typically to suppress their prey. However, if a predator is omnivorous, not only eating the prey but also consuming a more efficient predator on the same prey (i.e., it is an "intraguild predator"), it may actually relax the predation load on their shared prey, thus increasing the shared prey's abundance. For example, guilds of biological control agents must be carefully structured because some species eat not only the host but also other predators/parasitoids and thus their presence decreases the number of control agents and increases target pest populations. Many other cases exist in which consumers, via such intraguild predation, may indirectly facilitate its prey while concurrently exploiting it via direct consumption. Another example of indirect effects mediated by other than studied "focal species" is shown by the interaction between Australian bell miners and their homopteran food ("lerp"). After these birds were removed experimentally, the insects first increased greatly in number and then vanished when other bird species invaded the now undefended miner territories. Thus, the apparent effect of leaf miners on lerp insects (here, suppression or facilitation) depends on when the insects were surveyed. Such complications have undoubtedly interfered with clear interpretation of many experiments. The caveat is clear: Experiments can be indeterminate, producing contradictory, counterintuitive, or no results, depending on the relative strengths of the direct and indirect effects.

These problems can be anticipated and partially negated with the application of good intuition of the natural history of the system and important mechanisms. Such intuition is a product of intimate empirical knowledge gained through observation and guided by a conceptual awareness of which interactions are potentially important. Initially, this process is essential to design the appropriate experiments and identify which species and trophic links may be dynamically important. At the end, experiments must be interpreted in a food web context to assess possible indirect and higher order effects. Experimental results must be complemented with good descriptive, mechanistic, and comparative data to produce a deep understanding of the system. This is one role for energetic and dietary data. Experiments in the absence of natural history often do not succeed and may mislead.

The important messages from this section are that the complex food webs of natural communities can be simplified and understood by isolating key species and links into "interaction webs," experiments are absolutely necessary for this process, and experiments must be designed and interpreted with sound intuition based on natural history and theory.

III. OMNIVORY AND THE STRUCTURE OF FOOD WEBS

It is necessary to discuss feeding connections in more detail. Empirical research and logic have shown that the vast majority of consumers on this planet are very omnivorous, feeding on many types of food throughout the entire food web. This is not to say that all species are so catholic in their diets. Specialists abound, e.g., many herbivores or parasites consume only specific plants or hosts. However, these form a minority of consumers. The ubiquity of omnivory carries many implications for our efforts to produce theory and models to understand how food webs operate in and shape natural systems.

Omnivory occurs ubiquitously when consumers eat prey from general classes of prey, such as arthropods, plankton, soil fauna, benthos, or fish. The existence of multiple trophic types within these classes causes consumers to feed on species from many trophic levels. For example, "arthropodivores" eat whatever properly sized arthropods are available (e.g., predaceous spiders and insects and insect parasitoids, herbivores, and detritivores) without pausing to discriminate among their prey according to trophic status. For example, in the Coachella Valley desert delineated by Polis (1991) over 10 years of study, predaceous and parasitoid arthropods formed 41% of the diet of vertebrate and 51.5% of invertebrate arthropodivores, with the remainder of the diet being herbivorous and detritivorous prey. Similarly, inspection of diet data of planktivores, piscivores, "insectivores," carnivores, or benthic feeders reveals that such different channel omnivory is almost universal with the exception of those few taxa that specialize on a few species of prey.

Another important type of omnivory occurs when consumers eat whatever resources are available or abundant at a particular time or place, regardless of their trophic history. When analyzed, the diet of a single species usually shows great differences through time (e.g., seasonally) and space (patches or habitats). Prey exhibit three general phenologies: pulsed (population eruptions lasting a few days or weeks), seasonal (present for 2–4 months), and annual (available throughout the year). Feeding on prey from all three phenologies produces diet changes over time for almost all non-specialist consumers. Furthermore, many (most?) vertebrates opportunistically switch from plant to animal foods with season. For example, granivorous birds, rodents, and ants primarily eat seeds but normally feed on the abundant "arthropods" (= insects from all trophic levels and spiders) that appear during spring. Alternately, many omnivorous, arthropodivorous, and carnivorous species consume significant quantities of seed or fruit. In the Coachella Valley, 79% of 24 primary carnivores eat arthropods and/or plants; for example, coyotes eat mammals (herbivorous rabbits, rodents, and gophers; arthropodivorous antelope and ground squirrels; carnivorous kit foxes and other coyotes), birds (including eggs and nestlings, e.g., carnivorous roadrunners; herbivorous doves and quails), snakes, lizards, and young tortoises as well as scorpions, insects, and fruit. In New South Wales, 15 of 27 ant species are "unspecialized omnivores" eating nectar, seeds, plant parts, and a broad range of living and dead insects, worms, and crustacea. Overall, it appears that most consumers eat whatever is available and whatever they can catch.

"Life history" omnivory describes the great range of foods eaten during growth and ontogeny by most species (the "age structure component" of dietary niche breadth). Such omnivory includes abrupt diet changes in species undergoing metamorphosis (e.g., many marine invertebrates, amphibians, and holometabolic insects) and gradual diet changes in "slowly growing species" (e.g., reptiles, fish, arachnids, and hemimetabolic insects). Changes at metamorphosis can be great; for example, 22% of the insect families in the Coachella Valley desert community undergo radical change in diet—larvae are predators or parasitoids and adults are herbivores. Although not as dramatic, significant changes characterize slowly growing species so that differences in body size and resource use among age classes are often equivalent to or greater than differences among most biological species. Life history omnivory expands the diet of species throughout the entire animal kingdom with the exception of taxa that use the same food species throughout their lives (e.g., some herbivores) and those with exceptional parental investment (e.g., birds and mammals) so the young do not forage for themselves.

"Incidental omnivory" occurs when consumers eat foods in which other consumers live. Thus, scavengers and detritivores not only eat carrion or organic matter but also the trophically complex array of microbes and macroorganisms that live within these foods. Frugivores and granivores commonly eat insects associated with fruits and seeds. Predators eat not only their prey but also the array of parasites living within the prey. In each case, consumers automatically feed on at least two trophic levels.

These types of omnivory are widespread and common. Their ubiquity poses many questions. First, how does omnivory affect food web structure? Most obviously, it increases complexity and connectivity. Second, can we ignore omnivory in the analyses of food webs? By its very nature, omnivory causes consumers to have a great number of links, each of which may be numerically unimportant in the diet. For many reasons delineated later, we cannot arbitrarily ignore apparently minor diet links if we hope to understand dynamics.

IV. PATTERNS OF BIOMASS AND ENERGY IN FOOD WEBS

Primary productivity is among the most fundamental biological processes on the planet, transferring the energy locked in light and various inorganic molecules into forms useful to sustain producers and the diversity of consumers. What factors control primary productivity and regulate its distribution among plants, animals, and microbes? How do changes in primary productivity work their way through a food web to alter the abundance and biomass of herbivores to predators and detritivores? As discussed later, such key questions are best assessed using a food web approach. However, considerable controversy exists regarding the exact way that food web structure influences community and ecosystem dynamics.

A. Trophic Levels, Green Worlds, and Exploitative Ecosystems

Ecological research has amply demonstrated that food webs in nature contain hundreds to thousands of species, reticulately connected via multiple links of various strength to species in the autotroph and saprophagous channels and in the same and different habitats; omnivorous, age-structured consumers are common. Nevertheless, much food web theory still relies on the idealization of trophic levels connected in a single linear chain (plant herbivore carnivore). Here, we evaluate this simplification and some of its implications. In particular, we focus on two grand theories whereby food webs are considered to be central to community organization

The trophic level ideal in a simple linear food chain has had great appeal. Trophodynamics sought to explain the height of the trophic pyramid by reference to a progressive attenuation of energy passing up trophic levels, envisioned as distinct and functionally homogeneous sets of green plants, herbivores, primary carnivores, and, sometimes, secondary carnivores. This is a bottom-up community theory based on the thermodynamics of energy transfer. In counterpoint, Hairston, Smith, and Slobodkin's green world hypothesis (GWH; Hairston *et al.*, 1960) is primarily a top-down theory, with abundance at each level set, directly or indirectly, by consumers at the top of the chain. Thus, carnivores suppress herbivores, which releases green plants to flourish. These and earlier theoretical studies attempted to simplify food webs greatly to find generalities among them. GWH reduced complex webs to food chains in which species were pigeonholed into specific trophic levels. This allowed for predictions on how higher trophic levels (e.g., predators) influenced the dynamics of lower trophic levels (e.g., primary producers).

Oksanen *et al.*'s (1981) exploitation ecosystem hypothesis (EEH) generalizes GWH to fewer or more than three trophic levels. Trophic cascades are examples of food chains that behave approximately according to EEH. Trophodynamics and EEH each rely on the integrity of trophic levels and the existence of a single, albeit different, overwhelming mechanism that imposes structure on ecosystems. EEH proposes a conceptual framework of "exploitation ecosystems" in which strong consumption leads to alternation of high and low biomass between successive levels. Even numbers of "effective" trophic levels (two or four levels) produce a low-standing crop of plants because the herbivore population (level 2) flourishes. Odd numbers (one or three levels) result in the opposite effect: Herbivores are suppressed and plants do well. Proponents of EEH differ on subsidiary points, the first being the role of bottom-up effects in which primary productivity sets the number of effective levels. The most productive systems support secondary carnivores and therefore have four levels and low-standing crops of plants. Low-productivity systems (e.g., tundra) support only one effective level—plants. More productive habitats (e.g., forests) have three. Productivity is never high enough to support more than three effective levels on land or four in water. Other studies argue that physical differences between habitats, by affecting plant competition and consumer foraging, cause three levels on land and four in water.

EEH definitions of trophic levels are distinctive and adopt the convention that trophic levels occur only if consumers significantly control the dynamics or biomass of their food species. Without top-down control, consumers do not comprise an effective trophic level regardless of biomass or number of species involved. Supporters of EEH have noted that only when grazers regulate plants are grazers counted (as a trophic level), and only when predators regulate grazers are they fully counted. Thus, considerations of food chain dynamics do not become stranded in the immense complexity of real food webs. On the other hand, GWH trophic levels are based on energy deriving from primary productivity. Thus, "trophic level interactions ... weight particular links in the food web for their energetic significance." A trophic level is "a group of organisms acquiring a considerable majority of its energy from the adjacent level nearer the abiotic source." Despite these differences, both EEH and GWH theory argue that variability in the number of trophic levels exerts profound consequences on community structure and dynamics.

Considerable controversy exists as to the validity of GWH and EEH. The consensus has swung against these grand theories. Numerous arguments and empirical observations suggest that such processes operate occasionally in water but never on land. Basically, the complexity observed in natural systems does not conform to the reality of simple trophic levels. It appears that the notion that species clearly aggregate into discrete, homogeneous trophic levels is a fiction, arising from the need of the human mind to categorize. Especially in speciose systems, groups of species with diets of similar species do not occur. Omnivory, ontogenetic and environmentally induced diet shifts, and geographical and temporal diet heterogeneity all obscure discrete trophic levels. Even plants do not easily form a single level; higher plants have diverse crucial trophic and symbiotic connections with heterotrophs and many phytoplankton are mixotrophic, obtaining energy via photosynthesis, absorption of organic molecules, and ingestion of particles and bacteria. With increasing diversity and reticulation in webs, trophic levels blur into a trophic spectrum rather than a level. These species-individualistic and continuous "trophic spectra" are a reasonable alternative to the simplistic construct of homogeneous trophic levels.

B. Complex Food Webs, Multichannel Omnivory, and Community Structure

Polis and Strong (1996) offered a framework in the context of functioning community webs as an alternative to theories based on discrete trophic levels. Substantial evidence indicates that most webs are reticulate and species are highly interconnected, most consumers are omnivorous on foods (frequently on both plants and animals) across the trophic spectrum during their life history, most resources are eaten by many species across the trophic spectrum, plants are linked to a variety of species via trophic mutualism, most primary productivity becomes detritus directly, detrital biomass reenters the autotroph channel of the web when detritivores and/or their predators are eaten by consumers that also eat species in the herbivore channel, and species are often subsidized by food from other habitats.

They proposed that such trophic complexity pervades and generally underlies web dynamics. High connectance diffuses the direct effects of consumption and productivity throughout the trophic spectrum. Thus, consumer and resource dynamics affect and are affected by species at multiple positions along the trophic spectrum rather than interacting only with particular trophic levels. Consumer density is elevated and they often persist by eating resources whose abundance they do not influence (i.e., the interaction is "donor controlled").

Such dynamics are illustrated by focusing on top-down interactions. Some consumers exert "recipient" control on some resources and, occasionally, produce trophic cascades. Polis and Strong (1996) suggest that such control is often enabled by omnivorous feeding and various consumer subsidies that are usually donor controlled. Here, the transfer of energy and nutrition affects dynamics; numerical increases in consumer abundance occur from eating diverse resources across the trophic spectrum in the autotroph channel, from detritivores and detritus from the saprovore channel, from other habitats, and across their life history. Consumers, so augmented, exert recipient control to depress particular resources below levels set by the nutrition traveling through any particular consumer–resource link (analogous to the effects of apparent competition). Top-down effects arising from such donor-controlled, "multichannel" omnivory are depicted in Figs. 2 and 4. Strong consumer-mediated dynamics occur precisely because webs are reticulate and groups of species do not form homogenous, discrete entities.

Multichannel omnivory has two essential effects on the dynamics of consumers, resources, food webs, and communities. First, it diffuses the effects of consumption and productivity across the trophic spectrum rather than focusing them at particular trophic levels: It increases web connectance, shunts the flow of energy away from adjacent trophic compartments, alters predator–prey dynamics in ways contra to EEH assumptions, and thus disrupts or dampens the ecosystem control envisioned by EEH. For example, Lodge showed that omnivorous crayfish can depress both herbivorous snails (consistent with GWH and EEH) and macrophytes (inconsistent).

Second, omnivory can affect dynamics in a way analogous to apparent competition. Feeding on "nonnormal" prey can increase the size of consumer populations (or sustain them during poor periods), thus promoting top-down control and depression of "normal" prey. Frugivory, herbivory, granivory, detritivory, and even coprophagy form common subsidies for many predators. Vertebrate carnivores consume amply from the lower web without markedly depleting these resources. Does energy from fruit help carnivores depress vertebrate prey (e.g., herbivores)? Arthropodivory by seed-eating birds is the norm during breeding, with insect protein crucial to nestlings. Arthropodivory by granivores (and conversely, granivory by arthropodivores) must enhance bird populations and thus reduce seeds (arthropods) to a greater degree than if diets were not so augmented.

C. Trophic Cascades or Trickle

One prediction of GWH and EEH is that communities are structured by trophic cascades. Trophic experiments to test cascades use two methods: a bottom-up approach by increasing a resource (e.g., nitrogen or phosphorus) or a top-down approach that adds a top predator to a system. In the former, trophic cascades lead through a set of intermediate steps to increase densities of particular species or trophic groups higher in the web. In the latter, the top predator suppresses the trophic level below leading to increased densities two levels below. Thus, the expected responses should follow GWH/EEH predictions where alternating trophic levels are arranged with opposite densities (common—rare—common). For example, in a tritrophic (three-level) food chain, an increase in nutrients results in increases in the primary producer (plant) trophic level, decreases in the primary consumer (herbivore) level, and an increase in the top consumer level.

Proponents GWH and EEH suggest that strong trophic cascades occur in numerous food webs whereby entire trophic levels alternate in abundance via cascading food web interactions. However, empirical evidence shows that such cascades rarely or never occur on land and are apparently only present in a few aquatic communities. What determines whether a strong trophic cascade occurs or food web interactions weaken to become a trophic "trickle"? One major consideration is the efficiency of energy and resource transfer up the food chain. Highly efficient transfers lead to large numbers of top predators/consumers that would affect top-down control and strong cascades. Any factors that decrease the efficiency of energy/resource transfer would lessen the top-down control. In accordance with Polis and Strong's (1996) multichannel omnivory, an increasing list of factors have been examined to explain the differences between GWH/EEH expectations and experimental results and observations of natural communities that generally show weak or no trophic cascades. These factors include omnivory, ontogenetic shifts, edibility, food quality, ecological stoichiometry, cannibalism, disease, body size refuges (for prey), allochthonous resources, seasonality, life history characteristics, predator avoidance behavior, and spatial and temporal heterogeneity in the availability of resources.

V. CURRENT TOPICS/TRENDS IN FOOD WEB STUDIES

Here, relatively under-studied aspects of food webs perceived to be central to understanding populations, communities, and ecosystems are identified. Some of the topics are now focal points for food web research, both empirical and theoretical.

A. Food Webs as Open Systems

Recent methods of tracing stable isotopes through a food web can provide much information on feeding relationships and on the sources of productivity that drive communities. For example, using stable isotopes or diet data, one can determine whether a community utilizes resources that originate in the benthic or pelagic zones of lakes or both.

Virtually all natural systems are open and can exhibit tremendous spatial heterogeneity. Great spatial heterogeneity exists and nutrients and organisms ubiquitously move among habitats to exert substantial effects. However, food web studies have tended to focus on communities at a given site without regard to potential interactions with the surrounding habitat. Thus, little attention has been given to the fact that food web structure and dynamics are influenced by the movement of resources and organisms across habitat boundaries. Trophic linkage between habitats depends on the degree of differentiation in habitat structure and species composition. Systems that are moderately different tend to have broader transition zones and greatly overlap in species composition; these include grassland–forest, littoral–sublittoral, and benthic–pelagic zones. Habitats that have significant and abrupt changes in structure and species composition occur at the land–water interface.

Moving resources (energetic or nutrients) can be utilized by different trophic types and the organisms that move across boundaries may also differ trophically (e.g., predators and prey). Studies of communities on island systems have shown that most of the allochthonous inputs (i.e., input from other habitats) from the ocean are available to detritivores, predators, and scavengers. Such movement of nutrients, detritus, food, prey, and predators is absolutely ubiquitous, occurring in virtually all communities and across all habitats. Some systems heavily dependent on allochthonous inputs include caves; mountaintops; snowfields; recent volcanic areas; deserts; marine filter-feeding communities in currents; soil communities; the riparian, coastal areas; and lakes, rivers, and headwater streams that receive watershed inputs. However, all systems depend on allochthonous inputs. For example, recent work shows that plant productivity in both the Hawaiian Islands and the Amazon forest is dependent on phosphorus input from thousands of miles away (China and Africa, respectively). The migrations (e.g., songbirds or geese) and movement of herbivores (e.g., wildebeest or hippopotamuses) can also result in large energetic flows across habitats.

Allochthonous inputs into the top level include carrion or carcasses, the movement of prey species into the habitat, and movement of predators across habitats. For example, the Allen paradox describes cases in which secondary production within streams is insufficient to support levels of fish production in them. Similarly, studies of coyote populations along the coast in Baja California demonstrate that they are highly subsidized by inputs from the ocean (about half of their diet) and are able to maintain a 3 to more than 10 times higher density than in adjacent inland areas. Predators moving along the interface between ecosystems (i.e., shorelines, riverbanks, and benthic and pelagic systems) can utilize resources across habitat. The river continuum concept argues that allochthonous resources entering into small headwater streams provide much of the productivity for organisms downstream in larger order streams. These allochthonous resources include prey, dissolved and particulate organic matter, and litter fall. Such inputs also power estuarine systems in which rivers carry allochthonous inputs into estuaries. Similarly, runoff from terrestrial systems into aquatic systems (and vice versa) provides litter, dissolved and particulate organic matter, and prey.

Spatial coupling can be key to dynamics. For instance, arboreal anole populations, subsidized by insects imported from light gaps, increase so as to suppress some predators and herbivores. Abundant detrital kelp from the sublittoral zone promotes dense intertidal limpet and urchin populations that then graze noncoralline algae to low cover. Allochthonous subsidies commonly influence stream systems: Leaf fall subsidizes herbivores, which in turn depress algae. Spiders that live along the coasts of streams, rivers, lakes, or the ocean are often very dense because they feed on aquatic insects. These spiders can then depress herbivores and thus increase the success of plants on which they live. Such spatial subsidies appear to be the foundations of most of the well-known trophic cascades. All these interactions are donor-controlled: Consumers do not affect the rate of import, availability, or dynamics of the allochthonous resources. However, subsidies allow consumers to be more abundant than if supported solely by *in situ* resources, with consequent suppression of *in situ* resources decoupled from *in situ* productivity.

A common thread that has begun to link most thinking on food webs is that they are dynamical systems that vary over space and time. This approach has been liberating to ecologists, both empirical and theoretical. Recent empirical studies have found that communities and food webs contain multiple pathways that allow them to respond to environmental change and disturbance.

B. Detritus

Little of the energy fixed by plants passes directly into the grazing food chain—herbivores eating plants and then eaten by carnivores. Most of this primary productivity is uneaten by herbivores (median >80% on land, ~50% in water). What happens to this dominant chunk of the world's productivity? Is the detrital web a self-contained sink internally recycling energy and nutrients or a link that affects the population dynamics of the larger species?

Uneaten plants (and animals) enter the detrital web, in which they are processed by microbes, fungi, and some animals. Although some ecosystems are net accumulators of undigested biomass (e.g., carboniferous bogs and forests that supply today's oil and gasoline), most ecosystems do not accumulate plant biomass. Rather, it is soon digested by detritivores, with nutrients and energy passing through "functional compartments" composed of diverse microbes and animals. Several factors regulate the flow and availability of detritus to detritivores and then onto other consumers. A major question is rather whether the detrital community is a sink that metabolizes most of this energy or a link that passes this energy up the food chain.

An unknown fraction of detrital energy and nutrients re-enter grazing food chains when some detritivores are eaten by predators that also eat herbivores (e.g., a robin eats an earthworm. Such "detrital shunts" are common, interweaving energetics and dynamics of biophages and saprophages. Bypassing herbivores, this linkage can affect herbivore regulation in a manner analogous to the spatial subsidies to consumers discussed previously. Predator populations, subsidized by detritivorous prey, can increase and suppress other predators or herbivores.

The exact effect of detrital shunts depends on the relative benefits for each species and where detritus reenters (to producers, herbivores, and intermediate or higher consumers). For example, nutrients from detritus greatly influence plant productivity; models show that a 10% reduction in detritus can cause a 50% reduction of plant biomass. The dynamics of consumer control within the detrital web and those produced by infusion of detritivores into the grazing web are undoubtedly crucial to community structure and dynamics. For example, detrital shunts to predators in the grazing chain can create the appearance of a simple linear trophic cascade, but with the difference that nutrition from detritivores sustains or elevates predators to levels sufficient to suppress herbivores.

C. Age Structure Effects in Food Webs

Almost all species display complex life cycles, marked by moderate to radical changes in diet and habitat; such life histories fundamentally must affect every species with which they interact. However, our understanding of how age- and stage-structured processes affect food webs and communities is embryonic.

Life history omnivory describes shifts in diet during development; often, they are accompanied by ontogenetic changes in habitat. Diet can change substantially either discontinuously (e.g., at metamorphosis) or slowly with growth. Such life histories are widespread; an estimated 80% of all animal species undergo metamorphosis. Changes in resource use can be dramatic (e.g., predaceous juveniles, plant-feeding adults in parasitoids and many other insects, and herbivorous tadpoles and predaceous frogs and toads), with prey size variation as great as three or four orders of magnitude. Even among nonmetamorphic species, diets change greatly with age, with diet differences among age classes often more distinct than those among most species.

Overall, complex life histories and age structure omnivory can exert diverse and profound effects on the dynamics of populations and food webs. For example, they can either impede consumer control or amplify resource suppression via dynamics similar to those of spatial subsidy or detrital shunts.

D. The Roles of Nutrients and Stoichiometry

Animals require both energy and a variety of "nutritional requisites" to grow, complete their life cycle, and reproduce. Important nutrients include nitrogen, phosphorus, some trace elements, fatty acids, and vitamins. Nitrogen is an integral component of many essential compounds: It is a major part of amino acids, the building blocks of protein, including the enzymes that control virtually all cellular processes. Other nitrogen compounds include nucleic acids and chlorophyll. Phosphorus is used for adenosine triphosphate (ATP, the energy currency of all cells), nucleic acids (DNA and RNA), and phospholipids, particularly in cell membranes.

The availability of nutritional requisites constrains growth and reproduction in virtually every species. Nitrogen and phosphorus are particularly important. The ratio of carbon to nitrogen (C : N) in plants ranges from 10 : 1 to 30 : 1 in legumes and young green leaves to as high as 600 : 1 in some wood. The C : N ratios in animals and microbes are much lower, ordinarily between 5 : 1 and 10 : 1. Such differences in C : N ratios between plants and their consumers lower the rate of decomposition by microbes. There is ample evidence that heterotrophs chronically lack adequate nitrogen to grow or reproduce optimally. The importance of nutritional restriction is reinforced by the foraging literature that clearly shows that herbivores choose their foods based on nutrient as well as energy content.

In many cases, phosphorus availability constrains herbivore success. The Redfield ratio describes the approximate stoichiometric mix (110 C : 250 H : 75 O : 16 N : 1 P) of elements found in marine systems. In particular, the N : P ratio crucially determines productivity and species composition. Thus, energy (C–C bonds) and nitrogen could be abundant, but neither individuals nor populations grow maximally because phosphorus is insufficient. Because phosphorus is essential to cell division (and thus reproduction), a high N : P ratio especially limits the growth of organisms that have high potential $r_{\max}$, such as most herbivores and detritivores. These organisms are key to the potential regulation of plant biomass (and "detritus"). Evidence suggests that high N : P ratios can impede trophic cascades. For example, *Daphnia*, a key to many lake cascades, respond sufficiently rapidly to phytoplankton productivity to depress plant biomass. In lakes with inadequate phosphorus, slower growing copepods replace *Daphnia*; these copepods do not have the reproductive capacity to depress phytoplankton biomass.

Ecologists are beginning to understand how stoichiometry and nutritional balance affect population and food web dynamics. Nevertheless, it is extremely likely that herbivore growth is often less than maximal solely because their environment does not provide sufficient quantities of all key nutritional requisites. In fact, the greatest disparity in biochemical, elemental, and stoichiometric composition in the entire food web occurs at the link where herbivores convert plant material into animal tissue. The implication is clear: Even in a world full of green energy, many or most herbivores cannot obtain enough requisite resources to grow, survive, or reproduce at high rates. Nutritional shortages regulate herbivore numbers and often limit their effects on plant biomass.

Recent theoretical studies of the role of food quality in terms of edibility and nutrient content show that low food quality can greatly influence consumer resource interactions. This has two important consequences. First, low food quality reduces the growth rate of the consumer, making that interaction more stable. Second, in systems in which multiple resources could be limiting, the addition of large amounts of a single resource (such as nitrogen or phosphate) may increase that resource to a level at which it is no longer limiting; however, a second resource would become limiting and so on. This sequential limiting of resources means that the addition of a single resource would not push the system into highly unstable dynamics, reducing the probability that the "paradox of enrichment" occurs. Rosenzweig introduced the concept of the paradox of enrichment to explain the addition of a resource leading to the collapse of a consumer–resource interaction. This happens because the addition of the resource drives the population of the consumer to a higher level that results in overcompensation by the consumer (predator) driving the resource (prey) extinct. However, most systems have several potentially limiting resources. For example, Leibold's study of ponds found that nitrogen additions do not lead to strong trophic cascades or the paradox of enrichment because light becomes limiting with relatively modest nitrogen additions.

E. Interaction Strength

One goal of functional webs is the quantification of interaction strengths within food webs. Various definitions have been used for "interaction strengths." In Lotka–Volterra models, interaction strengths are due solely to the direct interactions between species pairs and are measured on a per capita basis. Estimations of the strength of these direct interactions are fraught with difficulties. Measurements in artificial systems may not allow for behavioral responses. For example, Sih has shown that prey species have different escape mechanisms or routes depending on the species of predator. Thus, when in the presence of two predators, the response of a prey may result in its increased susceptibility to one or the other predator due to a behavior that is not evidenced when only the one predator is present.

Measurements in natural systems are also problematic because they may not account for indirect interactions. Many studies have elucidated the interaction strength among pairs of species. However, indirect effects may play a strong role in determining the realized interaction strength. Thus, Paine has argued that interaction strengths should always be measured in the field with the full complement of natural species present and that these measurements should incorporate all indirect effects. The realized interaction strength accounts for all direct and indirect interactions. For example, predator–prey interactions are functionally negative due to the direct effect. However, the indirect effect of a predator may reduce the number of competitors of the prey species, thus resulting in an overall positive interaction strength (direct + indirect effects). Therefore, potentially strong indirect effects can make mechanistic interpretation of experimental results among species difficult.

Path analysis, a new statistical method, has been used to evaluate causal hypotheses concerning the strengths of interactions in many systems. Path analysis is essentially a multiple regression on each species in which specific causal relationships (e.g., alternative food web configurations), specific experimental treatments, and other interactions are diagrammed in a community interaction web. The community interaction is essentially a food web to which nonconsumptive interactions, such as pollination, competition, and mutualisms, are added. Hypotheses for the causal relationships between pairs of species not directly linked can become quite complicated. However, path analysis can test different hypothesized community web structures by accounting for both direct and indirect relationships. Then, experimental manipulations (e.g., species removals or additions) can test predictions of the path analysis.

F. Can Energetic Webs Provide Insight into Population and Community Dynamics?

A problem in food web studies is how to connect the great amount of quantitative information in energetic webs to population and community dynamics described by functional webs. Much progress would occur if we could determine the dynamical importance of a particular species or feeding link from an inspection of the magnitude of energy transfer or diet composition. Unfortunately, no clear answer is forthcoming. In fact, it appears that even highly quantified information such as the number of calories passed along a certain pathway or the frequency of prey in the diet of a consumer conveys little information about the dynamics of interacting populations because these descriptive parameters do not correlate with interaction strength.

There is no clear rationale to argue that food web dynamics and energetics are necessarily correlated; indeed, logic and evidence suggest dynamics often cannot be predicted from data on diet or energy flow. The degree of resource suppression is not a function of energy transfer. Consumer regulation of populations need involve little energy transfer and few feeding interactions. For example, removing predatory rats from New Zealand islands increased lizard abundance 3–30 times although lizards formed <3% of rat's diet. Key regulatory factors may produce much less overall mortality than other factors. Brief, intense predation episodes may net little energy for the predator but may be central to prey dynamics. The consumption of young stages (seeds, eggs, and larvae) may provide trivial energy to a consumer but can greatly depress prey abundance. Pathogens and parasites form an extreme example: They take little energy, even when they decimate their host populations. In a well-studied food web of the marine benthic community in the Antarctic, Dayton showed that the species apparently exerting the strongest effects on the structure and dynamics of this community would be deemed unimportant from analyses of diet, energy transfer, or biomass.

Such discoveries have stimulated many to argue that, without experimentation, one cannot a priori decide which are strong or weak links. An apparently weak link (in terms of diet or energy transfer) can be a key link dynamically, and an important energetic link may affect dynamics little. No necessary concordance of dynamics with either dietary or energetic measures exists. This insight counters the use of energetics to recognize strong interaction links.

G. Modeling Food Webs

To many ecologists, early food webs of Forbes, Summerhayes, and Elton and those of Lindeman emphasized the overwhelming complexity of natural systems and the need to simplify them into distinct trophic groups. This perspective was culminated in the greenworld hypothesis of Hairston *et al.* (1960). Oksanen *et al.*'s (1981) EEH expanded this view for ecosystems that had fewer or more than three trophic levels and for which the exact number of trophic levels was set by productivity. The top level would then regulate the one below it and this would release the one below it, etc. In this sense, both GWH and EEH suggested that all ecosystems are essentially regulated from the top-down by predation.

Lindeman envisioned the food web (or as he called it, the "food-cycle") as a dynamic system in which energy and nutrients are transferred from one trophic level to the next and recycled. This was an important departure from simply determining feeding connectedness (and from the GWH) in that ecosystems could be regulated from the bottom up by the flow of energy and materials from the level below. However, much more information and data are required to quantify the transfer of energy (and material) through food webs, but this view allows for a more analytical approach.

MacArthur focused the attention of ecologists on the trophic–dynamic approach with his hypothesis that increasing complexity of community organization leads to increasing dynamic stability. The reasoning was simple: When predators have alternative prey, their own numbers rely less on fluctuations in numbers of a particular species. Where energy can take more routes through a system, disruption of one pathway merely shunts more energy through another, and the overall flow continues uninterrupted.

MacArthur's analytical approach linked community stability to species diversity and food web complexity and it stimulated a flurry of theoretical, comparative, and experimental work. This work may be divided into two contemporary approaches that use food webs to study community structure. The first approach involves the study of the properties of food web diagrams with the goal of uncovering general patterns that suggest mechanisms of community stability. This is done both by comparing food webs from natural communities and by the use of simulation and mathematical modeling to study hypothetical food webs. This research has yielded much of the terminology now associated with food webs and generated a body of food web theory that includes many hypotheses about community structure.

The second approach, which grew from early theoretical and experimental community studies, involves the dynamical analysis of food webs to determine not only the pattern of interactions among the populations in the community but also the relative strengths of those interactions. Dynamic food web analysis also seeks to reveal interactions that are not obvious from simple food web diagrams, so-called indirect interactions. This approach requires the careful merging of experimental and theoretical approaches.

The simplicity of the GWH enabled it to be a reasonable starting point to examine the dynamics of food webs. In general, dynamical models are rooted in a tradition based on the application of Lotka–Volterra equations to communities and advocated by May (1973). One of the major conclusions from these phenomenological models [illegible] complexity (e.g., omnivory and long chains) [illegible] instability in model systems. This conclusion was viewed with skepticism by empiricists because observations from field studies (such as work by MacArthur) suggested that increased complexity should result in increased stability. Recent theoretical investigations into the relationship between stability and complexity have found that assumptions and structure of earlier models may have biased them toward decreased stability with increasing complexity.

Early theoretical studies of interactions and consequences of these interactions in food webs were based on equilibrium dynamics of Lotka–Volterra models. The assumption that ecological systems or species populations have some "equilibrium" around which they fluctuate is totally unrealistic. Furthermore, these early models ignored the central belief of many empiricists that most interactions between species were weak. The outcome of many of these theoretical studies went against common sense intuition and the findings of empirical studies, including that omnivory was destabilizing and therefore rare and that complexity (greater diversity) was also destabilizing. Recent studies that incorporated the findings of mostly weak interactions and nonequilibrium dynamics have found that omnivory and complexity may actually stabilize food webs. This agrees with both the intuition and the current arguments of empiricists who find that many weak interactions occur within food webs and these promote stability.

Recent theoretical studies suggested three factors as important to reduce stability in earlier models: (i) linear Lotka–Volterra equations, (ii) using equilibrium solutions to these equations, and (iii) the distribution of interaction strengths overly estimated the number of strong links. Many studies have shown that many predator–prey relationships are not linear, but instead predators exhibit saturation such as described by a Holling's type II functional response. Current models take advantage of this and use energetic uptake rates that saturate based on body size relationships. Also, equilibrium solutions to Lotka–Volterra relationships can give biologically unrealistic results because the assumption of equilibrium does not appear to hold in many predator–prey relationships. May and others used a uniform distribution in randomly created model food webs, which resulted in their webs having an overrepresentation of strong interaction compared to natural systems. This convention was based on the few early studies that examined the distribution of interaction strengths and suggested that there is a bias for weak interactions. May acknowledged that if the distribution of interaction strengths was not uniform, his results may not hold. Furthermore, recent theoretical studies also suggested that omnivory can stabilize food webs. Paradoxically, researchers using Lotka–Volterra models have found that although on average omnivory decreased stability, those systems in which omnivorous links persisted had the greatest stability. This increased stability may occur in Lotka–Volterra models when randomly created omnivorous links are weak.

In modeling food webs, a key consideration is the functional relationship between a consumer and its resource. As noted previously, Lotka–Volterra consumer–resource relationships are linear (type I). This assumes that the predators do not become saturated and can consume all available prey. Holling introduced nonlinear consumer–resource functional relationships with his disk (now called type II) functional response. This functional response assumes that the capture and consumption/digestion time of prey by the predator limits the amount of prey taken by a predator in a given amount of time. Holling also introduced a third type of functional response (type III) to simulate a predator switching capture of prey when a target prey species becomes rare to a more abundant prey species. Various other functional responses have been introduced, including Ginzburg and Arditi's ratio-dependent functional response. Ratio dependence assumes that the growth rate of the predator is dependent on the ratio of prey and predator densities, whereas in types I–III predator growth rates are dependent only on the prey densities (prey dependent). Ratio-dependent models predict that all trophic levels increase proportionately, whereas prey-dependent models predict the alternating pattern of GWH and EEH. The arguments against ratio dependence arise from the lack of a mechanistic basis for the model. Proponents of ratio dependence argue that the formulation accounts for the different timescales of reproduction and behavior. They also argue that the formulation is simpler and can account for essential dynamics of food webs without added complexity. Detractors argue that using more mechanistic, albeit more complex, models that can account for realistic interactions is the correct way to proceed. Regardless of these arguments, using intermediate levels of complexity based on realistic mechanisms is the current trend in food web theory.

H. Intermediate Levels of Complexity

Community ecology has focused on interactions (mainly competition and predation) between pairs of species that are fundamentally important in food webs.

However, these interactions, taken out of context of the larger web, may result in misleading information (due to indirect effects). Highly complex food webs, however, are unwieldy and intrinsically difficult to study in model systems. Thus, Holt and others suggested investigating the dynamics of intermediate (between species pairs and whole food webs) levels of complexity—so-called "community modules"—that are defined as small subsets of species that are characterized by strong interactions. These modules are also more representative of the levels of complexity (i.e., number of species) examined in experimental studies.

Recent theoretical studies have taken advantage of intermediate complexity by focusing on food web interactions that typify interactions found in real food webs but are common to many food webs. This allows one to examine how indirect effects interact with direct effects to structure food webs. Common types of interactions among sets of species and their resources (modules) are apparent competition, intraguild predation, omnivory, cannibalism, and spatial subsidies.

This modular approach has allowed for various theoretical studies to examine stability of food webs using mathematical approaches. For example, McCann, Hastings, and Huxel found that adding relatively weak interactions among species could enhance food web stability. They found this to be true for apparent competition, intraguild predation, omnivory, cannibalism, and spatial subsidies.

Are weak interactions typical of food webs? The answer, from the few studies that have specifically examined this question, is yes. For example, in a study on intertidal food webs, Paine found that most interaction strengths are weak. Moreover, knowing that most predators eat tens to ~100 species of prey suggests that most of these interactions are weak.

One may then ask, what about strong interactions? The answer goes to the heart of one major problem with earlier food web studies. Strong interactions may occur and be a regular component of food webs. However, in almost every case, they appear to be enabled by "multichannel omnivory" (i.e., feeding on many weak links; Polis and Strong, 1996) or are restricted temporally and/or spatially because they are inherently unstable. However, time and effort constraints and tradition have caused the vast majority of food web studies to ignore the many weak interactions and the spatial and temporal aspects that characterize all systems.

In another theoretical study of food web processes that took advantage of the modular view, Huxel and McCann examined the flow of the allochthonous energetic resources. They found that allochthonous resources may spread evenly throughout the community or may become compartmentalized. High levels of allochthonous resources decreased stability, whereas low levels increased stability. Thus, again a weak link tended to increase stability.

See Also the Following Articles

DIVERSITY, COMMUNITY/REGIONAL LEVEL • ECOSYSTEM, CONCEPT OF • ENERGY FLOW AND ECOSYSTEMS • KEYSTONE SPECIES • PREDATORS, ECOLOGICAL ROLE OF • SPECIES INTERACTIONS • TROPHIC LEVELS

Bibliography

DeAngelis, D. L. (1992). *Dynamics of Nutrient Cycling and Food Webs.* Chapman & Hall, New York.

Forbes, S. (1887). The lake as a microcosm. *Bull. Illinois State Nat. History Surv.* 15, 537.

Hairston, N. G., Sr., Smith, F., and Slobodkin, L. (1960). Community structure, population control and competition. *Am. Nat.* 94, 421.

May, R. (1973). *Stability and Complexity in Model Ecosystems.* Princeton Univ. Press, Princeton, NJ.

Oksanen, L., Fretwell, S., Arruda, J., and Niemela, P. (1981). Exploitation ecosystems in gradients of primary production. *Am. Nat.* **118**, 240.

Paine, R. T. (1980). Food webs: Linkage, interaction strength and community infrastructure. *J. Anim. Ecol.* 49, 667.

Polis, G. A. (1991). Complex trophic interactions in deserts: An empirical critique of food web theory. *Am. Nat.* 138, 123.

Polis, G. A., and Strong, D. R. (1996). Food web complexity and community dynamics. *Am. Nat.* 147, 813.

Yodzis, P. (1989). *Introduction to Theoretical Ecology.* Harper & Row, New York.

529
2005
0

FOREST CANOPIES, ANIMAL DIVERSITY

Terry L. Erwin
Smithsonian Institution

GLOSSARY

arbicolous Living on the trees, or at least off the ground in shrubs and/or on tree trunks.

emergent A very tall tree that emerges above the general level of the forest canopy.

epiphytic material Live and dead canopy vascular and nonvascular plants, associated detritus, microbes, invertebrates, fungi, and crown humus.

hectare Metric equivalent of 2.47 acres.

microhabitat A small self-contained environmental unit occupied by a specific subset of interacting species of the forest (or any other community).

scansorial Using both the forest floor and canopy for movement and seeking resources.

terra firme forest Continuous hardwood forest of the nonflooded or upland parts of the Amazon rain forest.

THE FOREST CANOPY is arguably the most species-rich environment on the planet and hence was termed the "last biotic frontier," mainly because until very recently it had been studied less than any place else, with the exception of the deep ocean floor and outer space. The reason for lack of study of the canopy was accessibility, and the evidence of the incredible species richness, mainly of tropical forests, is primarily the abundance of insects and their allies. This hyperdiverse and globally dominant group has adapted to every conceivable niche in the fine-grained physical and chemical architecture of the tree crowns. In less than three decades, canopy biology has become a mixed scientific discipline in its own right that is gradually gaining sophistication of both approach and access.

Tropical arbicolous (tree-living) arthropods were observed in the early 1800s in the "great forests near the equator in South America" and later that century were described by Henry Walter Bates. Even though Bates observed, described, and commented on the canopy fauna (as viewed from the ground and in recently felled trees), more than a century passed before Collyer designed an insecticide application technique that allowed a rigorous sampling regime for canopy arthropods. William Beebe and collaborators early in the twentieth century recognized that the canopy held biological treasures, but "gravitation and tree-trunks swarming with terrible ants" kept them at bay. Frank Chapman, a canopy pioneer (of sorts), viewed the treetops from his "tropical air castle" in Panama in the 1920s, but his interest was vertebrate oriented, his perch was a small tower, and his observations of insects and their relatives were casual. By the mid-1960s and early 1970s, a few workers in both basic and applied science were seriously

Encyclopedia of Biodiversity, Volume 3

investigating canopy faunas of temperate and tropical forests in both the Western and Eastern Hemispheres. From the early 1980s until now, many workers have been improving methods of access and other techniques used to register, sample, and study the fauna (see reviews by Basset, Erwin, Malcolm, Moffett and Lowman, Munn and Loiselle, and Winchester in Lowman and Nadkarni, 1995; Moffett, 1993; Mitchell, 1987). Some of these workers have found that arthropods by far make up the fauna of the canopy (Erwin, 1982, 1988). Visiting and nesting bird, mammals, reptiles, and amphibians represent a mere 1% or less of the species and even less in the abundance of individuals in these groups (Robinson, 1986). There are no adequate measures of canopy nematodes, mollusks, or other nonarthropod microfauna groups.

What is meant by the forest canopy? Generally, the canopy, or tree crown, is thought of as that part of the tree including and above its first major lateral branches. The canopy of a single tree includes the crown rim (the leaves and small twigs that face the main insolation from the sun) and the crown interior (the main trunk and branches that gives a tree its characteristic shape). The canopy fauna is that component of animal life that inhabits the tree canopy and uses resources found there, such as food, nesting sites, transit routes, or hiding places. Hence, the forest canopy is collectively all the crowns of all the trees in an area. The canopy is often thought of as being stratified into emergents, one to three regular canopy strata, and an understory of smaller trees living in the shade of a more or less continuous overstory. All types of forests have their own describable characteristics, from the spruce forests of the Northwest Territories of Canada to the pine forest of Honduras, the dry forests of Costa Rica and Bolivia, and the Rinorea and Mauritia forests of the upper Manu River in Peru. It is through "whose eyes" one views the community, habitat, or microhabitat that determines the scale of investigation and subsequent contribution to the understanding of the environment—the beetles, the rats, the birds, the ocelots, the investigators, or perhaps even the trees.

I. CANOPY ARCHITECTURE, ANIMAL SUBSTRATE

A temperate forest is composed of both broad-leaved and coniferous trees, with one or the other sometimes occurring in near pure stands depending on the latitude and/or altitude and also on soil and drainage conditions. Normally, there are few canopy vines or epiphytes and perhaps some wild grape or poison ivy vines. Soil and organic debris caches are few or absent in the tree crowns, except for tree holes which provide homes to numerous arthropod groups but few vertebrates. Temperate forests are subjected to cold and hot seasonal climate regimes as well as wet and dry periods. Great expanses of forest lose their leaves in the winter months, sap ceases its flow, and the forest "metabolism" comes to a slow resting state.

The temperate forest seemingly provides a great variety of substrates for the canopy fauna, but faunas are depauperate compared to those in tropical forests. Virtually no mammals are restricted to temperate forest canopies—only a few frogs and lizards. However, many bird species are restricted to the canopies, as they are in tropical forests. Among insects, for example, the beetle family Carabidae has 9% of its species living arboricolously in Maryland, 49% in Panama, and 60% or more at the equator in South America.

Tropical forests, on the other hand, have few if any coniferous trees; only forests at higher elevations and/or located closer to subtropical zones have coniferous trees. Tropical canopies are often (but not always) replete with vines and epiphytes, tree holes, and tank bromeliads, and there are soil mats among the roots of orchids, bromeliads, and aroid plants. In the early 1990s, Nadkarni and Longino demonstrated that epiphytic material is fraught with macroinvertebrates, and Coxson and Nadkarni later showed that epiphytic material is important in the acquisition, storage, and release of nutrients.

Lowland tropical forests are subjected to mild temperatures, without frost, but have both wet (sometimes severe) and dry seasons. Individual species of trees may be deciduous, but in general tropical forests are always green and there is a perpetual growing season. Substrates are constantly available for the fauna. Often, some microhabitats with their substrates are temporary in the sense that they remain in place for a season or two, but then their architectural structure collapses into a jumbled pile of organic detritus on the forest floor. Such microhabitats (e.g., a suspended fallen branch with its withering leaves) provide a home resource to thousands of arthropods in hundreds of species, many found only in this setting. Eventually, such a branch loses its dried leaves and crashes to the forest floor. However, a short distance away, another branch breaks from a standing tree and the process begins again. The arthropods of the old, disintegrating branch move to the new one. The microhabitat and its substrates are forever present across the forest; each individual branch is ephemeral. The faunal members occupying such microhabitats are good at short-range dispersal.

II. EXPLORING THE LAST BIOTIC FRONTIER

Until recently, the forest canopy was impossible to study well. Getting there was the limiting factor, and even after getting there (e.g., via ropes) it was difficult to find the target organisms. Modern devises such as aerial walkways (e.g., ACEER, Tiputini Biodiversity Station; Fig. 1), one- or two-person gondolas maneuvered along crane booms (e.g., in Panama at STRI), and web-roping techniques (see review by Moffett and Lowman in Lowman and Nadkarni, 1995) now allow real-time observations, sampling, and experiments anywhere in the canopy. Inflatable rafts that suspend mesh platforms resting on the upper crown rims of several trees have provided access from above, although this technique seems more suited to botanical work or leaf-mining insects, especially epiphytes and lianas. Insecticidal fogging techniques allow passive sampling of all arthropods resting on the surfaces of canopy plants (Erwin, 1995), and suspended window/malaise traps collect the active aerial fauna. Many of these techniques have been used during the past two decades; however, often they were simply used as collecting devises to garner specimens for museums and/or for taxonomic studies, and for this purpose they are excellent. In some cases, ecological studies were desired, but the techniques were not properly applied and the results disappointing. It is important to first ask the questions and then design the experiments; in some cases, current canopy techniques can be powerful tools for answering questions. Unfortunately, although sampling is relatively easy, sample processing is time-consuming and laborious. For canopy fogging studies, after the sampling effort an average of 5 years was required before published products were achieved (Erwin, 1995). The main reason for this is a lack of funding for processing the results of fieldwork, even though the field studies were readily funded. Without processing, the data inherent for each specimen are unavailable for taxonomy or ecology studies. This is an historical funding problem and one of the reasons most studies examine but a few species from few samples.

III. RESULTS OF STUDIES

A. Invertebrates

Recent findings by Adis in the central Amazon Basin and by Erwin in the western part of the basin demonstrated that there are as many as 6.4×10^{12} terrestrial arthropods per hectare. A recent 3-year study of virgin

FIGURE 1 The rainforest canopy of the western Amazon Basin from the canopy walkway of the ACEER Biological Station.

terra firme forest near Yasuni National Park in Ecuador by Erwin found an estimated 60,000 species per hectare in the canopy alone. This figure was determined by counting the actual species in the samples of several well-known groups and comparing their proportions in the samples with their known described taxonomic diversity. The predatory beetle genus, *Agra* (Fig. 2), has more than 2000 species found only in Neotropical forest canopies and scattered remnants of subtropical forest canopies in southern Texas and northern Argentina. The herbivorous weevil genus, *Apion*, likely has more than 10,000 species. In only 100 9-m^2 samples of canopy column from 1 ha of virgin terra firme forest near Yasuni National Park in Ecuador, there are more than 700 species of the homopteran family, Membracidae, which were found along with 308 species of the beetle family, Carabidae, and 178 species of the spider family, Theridiidae.

"Biodiversity" by any other name is "Terrestrial arthropods"—that is, insects, spiders, mites, centipedes, millipedes, and their lessor known allies.

Forest canopy studies of terrestrial arthropods are few (Erwin, 1995). Many of these studies currently concentrate on host specificity as a herbivore or parasite that eats only one other species of plant or animal. However, there is another class of specificity that is very important in understanding biodiversity that has received almost no study: "where" species hide and rest. This is not random but rather species specific (T. L. Erwin, unpublished data).

Terrestrial arthropods are found in "hotels" and "restaurants" or "in transit" between the two (Fig. 3). Often, insects and their allies eat, mate, and oviposit in the restaurant or at the food source, for example, on fungi or in suspended dry palm fronds. These insects may hide during the day under debris or under bark near the fungus or on the palm debris, but they never roam far from the vicinity of the food source, except to locate new food sources when the old one is depleted. Members of other species eat in one place and then move to cover for a resting period, i.e., the hotel. An example of this is the subfamily Alleculinae of the beetle family Tenebrionidae. These beetles feed on lichens and moss on tree trunks at night and spend the day (hiding, resting, and possibly sleeping) in suspended dry leaves elsewhere in the forest. Many species found in the forest canopy during the day (utilizing leaves, fruits, and/or flowers) hide and rest at night in the understory (e.g., various pollen-feeding beetles and the larger butterflies).

Insects particularly, and some of their allies, have adapted to nearly every physical feature of the planet, and the canopy is no exception. Many beetles have special feet for walking on leaves; some even have modified setae on their feet to slow them down upon landing from rapid flight (Fig. 4). Because they are in an environment with raptorious birds, lizards, and frogs, many insect species have evolved camouflage coloration.

Climate is the main constraint on terrestrial invertebrates. In the temperate zones, it is the winter cold and

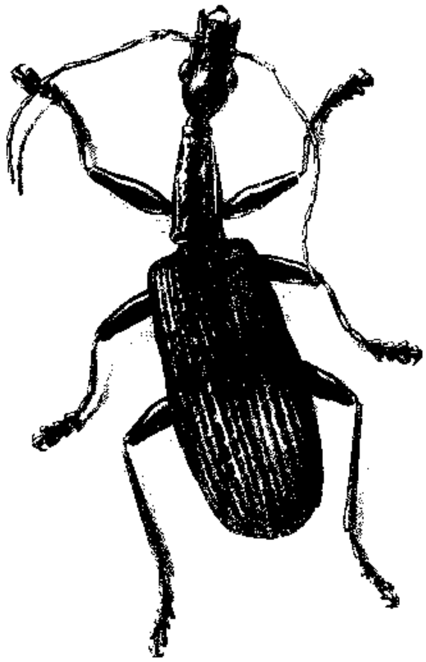


FIGURE 2 *Agra eowilsoni* Erwin, a species of Colombia, South America.

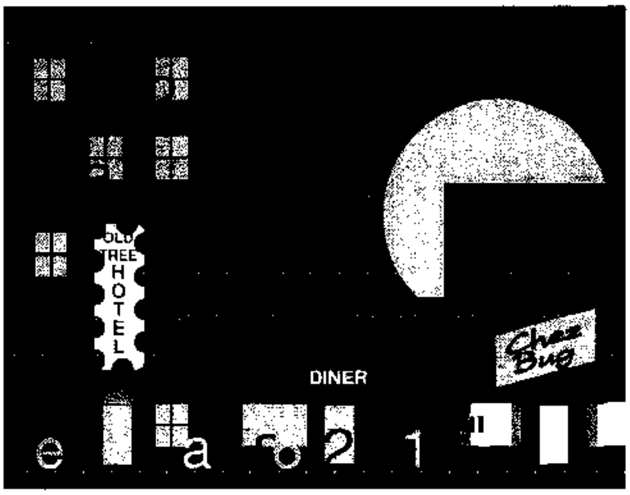


FIGURE 3 Humorous depiction of where "bugs" live and eat.

FIGURE 4 Setae of an arboreal beetle's tarsi used for landing and stopping quickly.

dryness; in the equatorial tropics, it is the dry season for some and the rainy season for others, with the temperature far less of an influence than it is in the far north or south. Many herbivores must contend with plants that produce toxic chemicals or other defensive systems. All insects must also deal with other insects that predate, parasitize, or carry bacteria, fungi, or other insect diseases. Hammond, Stork, and others, in their studies of insects in the Sulawesi dipterocarp forests, and Miller, Basset, and others in New Guinea found much less insect diversity and richness than Erwin and his teams in the Neotropical forests. Hammond also found in southwest Asia that the canopy fauna was not as delimited from the understory fauna as it is in the Amazon Basin. Unfortunately, all these teams used different methodology; hence, much of their results are not comparable. It is certain, however, that the Old World tropical forests are not as biodiverse as those in the New World, nor are the forests of Costa Rica and Panama as diverse as those of the Amazon Basin. Disparate regional richness is one of the main problems in estimating the number of species on the planet. Another is the incredible richness of terrestrial arthropod species and the fact that scientists likely know less than 3–5% of them if published estimates of 30–50 million extant species are close to reality. Stork (1988) has even gone so far as to suggest that there could be 80 million species on the planet.

B. Vertebrates

Availability of food year-round constrains vertebrates from living strictly in canopies (see reviews by Emmons and Malcolm in Lowman and Nadkarni, 1995). Only in evergreen rain forests is there a continuous supply of food (albeit somewhat dispersed and sporadic) for phytophagous and insectivorous vertebrates. In deciduous forests, most species also forage on the ground or hibernate when food supplies are short. Almost all canopy mammals live in evergreen tropical forests, but even there most are scansorial. Timing and distribution of food resources are the critical controlling factors. Among all nonflying vertebrates, anurans and lizards and to a lesser extent snakes are the most important truly canopy creatures. Birds and bats are also exceedingly important components. All these groups except snakes account for vertebrate predator-driven evolution on the far more dominant invertebrates of the canopy. For example (as Blake, Karr, Robinson, Servat, Terbourg, and others have shown), throughout the tropics approximately 50% of birds are strictly insectivores, whereas another 8% take insects and nectar.

Morphological adaptations that allow canopy life include feet that can firmly grip the finely architectured substrate of twigs, leaves, and scaly bark. Emmons, in her many articles on Neotropical mammals, demonstrated that among these animals, those with the ability to "jump" avoided wasting energy and time by descending and climbing new trees to find resources; hence, more true canopy species have this ability. This is certainly true also of frogs and lizards. However, it is the flying forms—birds and, to a lesser extent, bats—that account for most of the treetop vertebrate fauna. Physiological adaptations that allow vertebrate canopy life include the ability to subsist on diets of fruit, flowers, leaves, or insects and their allies. Among mammals, fruit eaters are dominant.

As shown by Duellman, Dial, and others, among canopy anurans and lizards, nearly all are primarily insect predators. Birds are overwhelming insectivorous in the canopy fauna, with approximately 40% in the upper Amazonian and 48% at Costa Rica's La Selva

Biological Station. Malcolm, in summarizing the few articles on the subject, estimates that 15% of mammal species are arboreal/scansorial in temperate woodlands, whereas between 45 and 61% exhibit this behavior in tropical forests. In Duellman's 1990 list of anurans and reptiles from Neotropical forest, 36% are strictly arbicolous, whereas 8% are scansorial. Among birds, Blake and others found that scansorial species using the understory and ground were more numerous than strictly canopy species (51 and 42%, respectively), at their site in Costa Rica.

In summary, although canopy vertebrates are important in driving part of invertebrate evolution in the forest canopy, they have not overwhelmingly radiated into or made use of the canopy, as have the invertebrates. For example, the total vertebrate fauna known at Cocha Cashu, Peru, is approximately 800 species (approximately 45% of which are arbicolous or scansorial), whereas at a nearby location there are nearly 900 species of the beetle family Carabidae, of which more than 50% are strictly arbicolous. In Ecuador, near Yasuni National Park, there are in excess of 600 species of the homopteran family Membracidae in a single hectare, 100% of which are strictly arbicolous.

IV. CONCLUSIONS

Although animals may use the air for dispersal, they live on substrate. Here, they eat, mate, hide, and walk. Forest canopies are rich in species because they offer a three-dimensional array of varying substrates that directly receive the sun's energy with little filtering.

Although much has been and is being accomplished by faunal studies of the forest canopy, there is still much to do. There are missing data links between vertebrates and invertebrates and between both of these and the plant food and plant architecture on which they depend, and data is also missing on the influence of the canopy physical features on the fauna such as microclimates (see Parker's review in Lowman and Nadkarni, 1995). Each subsystem is receiving at least some attention, but the new discipline of canopy biology is in its infancy. Is it too late? The forests and their species-rich canopies are rapidly disappearing (World Resources Institute, 1993).

Topics of current investigation include canopy insect β diversity and measures of host specificity, the latter particularly in leaf-feeding beetles. Both areas of study were driven by earlier, somewhat naïve estimates of millions of species extant on the planet (Erwin, 1982; Stork, 1988; May, 1990; Casson and Hodkinson, 1991; Gaston, 1991). Although some of these studies may have been internally consistent within the parameters set for the estimations, no one had really gotten a handle on the true meaning of "host" specificity, biocomplexity of tropical forests, the influence of tropical biotope mosaics, β diversity or what is known as species turnover in space and/or time, or the disparities of richness among continents or even the disparity among regions within continents.

Even so, our current rudimentary knowledge indicates that we are losing hundreds, even thousands, of invertebrate species with "scorched earth" programs such as that in Rondonia, Brazil, clear-cutting of Borneo and other southern Asian forests, and other losses in Haiti, Puerto Rico, Hawaii, the western Amazon Basin, Madagascar, and so on.

Conservation strategies are currently dominated by data on vertebrates (Kremen *et al.*, 1993; Samways, 1994); however, invertebrates are rapidly becoming sufficiently known to include them in analyses that are directed toward preservation of forest communities; to this end, the collective human conscience will soon be dealing with real extinction processes equivalent to those in the past, from the Permian to the Cretaceous. We are living at the beginning of the so-called "sixth extinction crisis" sensu Niles Eldridge of the American Museum of Natural History. Amelioration of the impact of this crisis rests on a better knowledge of the natural world around us and the development of conservation strategies that consider what we, Earth's managers (whether we like it or not), want future evolution to look like, as so well described by David Quammen (1998).

See Also the Following Articles

AMAZON ECOSYSTEMS • ARTHROPODS, AMAZONIAN • BEETLES • FOREST CANOPIES, PLANT DIVERSITY • FOREST ECOLOGY • INVERTEBRATES, TERRESTRIAL, OVERVIEW • TROPICAL ECOSYSTEMS

Bibliography

Casson, D. S., and Hodkinson, I. D. (1991). The Hemiptera (Insecta) communities of tropical rain forests in Sulawesi. *Zool. J. Linnean Soc.* **102**, 253–275.

Erwin, T. L. (1982). Tropical forests: Their richness in Coleoptera and other Arthropod species. *Coleopterists Bull.* 36, 74–75.

Erwin, T. L. (1988). The tropical forest canopy: The heart of biotic diversity. In *Biodiversity* (E. O. Wilson, Ed.), pp. 123–129. National Academy Press, Washington, D.C.

Erwin, T. L. (1995). Measuring arthropod biodiversity in the tropical forest canopy. In *Forest Canopies* (M. D. Lowman, and N. M. Nadkarni, Eds.), pp. 109–127. Academic Press, San Diego.

Gaston, K. J. (1991). The magnitude of global insect species richness. *Conserv. Biol.* 5, 283–296.

Kremen, C., Colwell, R. K., Erwin, T. L., Murphy, D. D., Noss, R. F., and Sanjayan, M. (1993). Terrestrial arthropod assemblages: Their use in conservation planning. *Conserv. Biol.* 7(4), 796–808.

Lowman, M. D., and Nadkarni, N. M. (1995). *Forest Canopies*. Academic Press, San Diego.

May, R. M. (1990). How many species? *Philos. Trans. R. Soc. London Ser. B* 330, 293–304.

Mitchell, A. (1987). *The Enchanted Canopy: Secrets from the Rainforest Roof*. Fontana/Collins, London.

Moffett, M. (1993). *The High Frontier—Exploring the Tropical Rainforest Canopy*. Harvard Univ. Press, Cambridge, MA.

Quammen, D. (1998, October). Planet of weeds, tallying the losses of Earth's animals and plants. *Harpers Magazine*, 57–69.

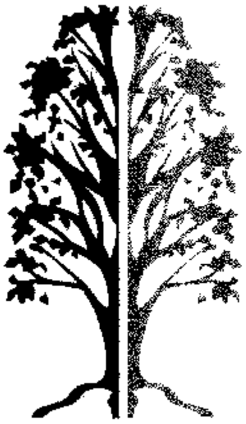
Robinson, M. H. (1986). The fate of the tropics and the fate of man. *Zoogoer* 5, 4–10.

Samways, M. J. (1994). *Insect Conservation Biology*. Chapman & Hall, London.

Stork, N. (1988). Insect diversity: Fact, fiction, or speculation. *Biol. J. Linnean Soc.* 35, 321–327.

Stork, N. E., Adis, J., and Didham, R. K. (1997). *Canopy Arthropods*. Chapman & Hall, London.

World Resources Institute (1993). World Resources 1992–1993. World Resources Institute/Oxford Univ. Press, New York.



FOREST CANOPIES, PLANT DIVERSITY

Nalini M. Nadkarni,* Mark C. Merwin,* and Jurgen Nieder†
*The Evergreen State College and †Botanisches Institüt der Universität Bonn

GLOSSARY

accidental epiphyte Plant that normally grows terrestrially but that occasionally grows to maturity in a tree crown, usually in terrestriallike microsites such as the crotches of branches.

bryophyte Nonvascular plant of the division Bryophyta (a moss, liverwort, or hornwort).

cryptogam Plant that reproduces by spores or gametes rather than seeds; includes bryophytes and lichens.

epiphyll (folicolous) Plant that grows on the leaf surface of another plant.

epiphyte Nonparasitic plant that uses another plant as mechanical support but does not derive nutrients or water from its host.

facultative epiphyte Plant or lichen that commonly grows epiphytically and terrestrially, usually exhibiting preference for one or the other habit in a particular habitat.

homoiohydry Ability to maintain a constant internal water balance independent of fluctuating environmental conditions.

lichen Composite organism consisting of a fungus (the mycobiont) and an alga and/or a cyanobacteria (the phycobiont) that live in a symbiotic relationship.

mistletoe Woody parasite that taps the xylem of a tree, but is capable of photosynthesis.

obligate epiphyte Plant that always grows on another plant for structural support, but derives no nutrients from the host.

parasite Woody or nonwoody plant that taps into the vascular system of a host plant and derives energy and/or nutrients from it, often to the detriment of the host.

poikilohydry Condition of internal water balance varying with changes in ambient humidity.

primary hemiepiphyte Plant that begins its life cycle anchored in a tree crown and ultimately becomes rooted in the ground (e.g., strangler fig).

secondary hemiepiphyte Plant that begins its life cycle as a terrestrial seedling, ascends a tree, and can later lose root connections with the ground, including (a) lianas, woody climbing plants with relatively thick stems that generally grow in mature habitats, and (b) vines, herbaceous climbing plants that regularly grow in disturbed habitats or forest edges.

THE IMPORTANCE OF THE PLANTS THAT DWELL IN FOREST CANOPIES is becoming increasingly recognized in relation to understanding biodiversity. The

Encyclopedia of Biodiversity, Volume 3

upper tree canopy of many forest ecosystems fosters extremely diverse plant communities, which include vascular and nonvascular epiphytes, hemiepiphytes, and parasites. Canopy-dwelling plants contribute substantially to overall forest biodiversity and biocomplexity by providing resources for arboreal vertebrates, invertebrates, and microbes, and by participating in nutrient and water cycling and gas and energy exchanges.

I. INTRODUCTION

A. Definition of the Forest Canopy

The forest canopy has been called "the last biotic frontier" (Erwin, 1988). It presents a habitat conducive to the evolution of literally thousands—perhaps millions—of species of plants, microorganisms, insects, birds, and mammals that are rarely or never encountered on the forest floor. Although forest canopies have been among the most poorly understood regions of our planet, their mysteries are being explored by increasing numbers of biologists. Canopy communities are now believed to be important in maintaining the diversity, resiliency, and functioning of the forests they inhabit.

The forest canopy is a structurally complex and ecologically important subsystem of the forest. It is defined as "the aggregate of all crowns in a stand of vegetation, which is the combination of all foliage, twigs, fine branches, epiphytes as well as the interstices (air) in a forest" (Parker, 1995). The forest canopy is the primary site of gas exchange between the atmosphere and vegetation and fosters many ecosystem processes that are crucial to the maintenance and diversity of the forest as a whole (Lowman and Nadkarni, 1995).

B. Scope of This Article

Trees are the most obvious structural component of forest canopies. Their trunks, branches, and leaves constitute the infrastructure of the canopy and provide mechanical support for thousands of species of arboreal plants and animals. Tree species diversity is discussed elsewhere.

Although much has been published on canopy plants, the question of the global importance of epiphytes for the biodiversity of tropical forests in general and the canopy in particular has not been explicitly addressed. Previous reviews of canopy biodiversity have primarily dealt with arthropods (Erwin, 1988), or with specific subgroups of canopy plants (Kress, 1986; Rhoades, 1995). Here, we describe the diversity of many types of canopy-dwelling plants. Included in this review are epiphytic vascular plants, epiphytic cryptogams (nonvascular plants that include lichens and bryophytes), primary and secondary hemiepiphytic vascular plants (lianas and vines), and arboreal parasitic vascular plants. Arboreal fungi and free-living algae are so poorly known that there is little to review. We place greatest emphasis on obligate and facultative epiphytes, and exclude "accidental epiphytes" from this review.

The term "diversity" in the following will be based on species as the unit of biological diversity, since assessment of other aspects of biodiversity is virtually nonexistent in the case of canopy plants. We first review the systematic distribution of canopy taxa and provide species counts based on the state of current knowledge. We discuss gradients of canopy plant diversity of microsites within the canopy at various spatial scales, spanning a single microsite within a tree (e.g., twig, branch bifurcation) to regional and global levels. Biogeographical analyses of canopy-dwelling taxa are then considered, as well as some of the major evolutionary elements that have influenced their distribution and abundance. "Habit diversity" (the diversity of morphological and physiological features) of arboreal taxa will then be described. Finally, we discuss conservation efforts that involve canopy plants and suggest future research possibilities.

C. Historical Roots and Sources of Information

In 1832, Charles Darwin first described what he termed the great diversity and profusion of "parasitical plants" (that we now understand to have been epiphytes), which he encountered in abundance in the coastal forests of Brazil. In the late nineteenth century, the German botanist A. F. W. Schimper first described epiphytes and outlined their importance to tropical botany.

Historically, canopy studies have been dominated by people who sought the thrill of climbing and followed the lure of discovering new species. Early European explorers hired climbers and trained monkeys to collect specimens of "exotic" air-plants that grew out of reach. Pioneering work in old-growth forests of the Pacific Northwest contributed to the application of mountain-climbing techniques for safe and reliable access to the canopies of tall trees. Since 1980, the innovation of high-strength and low-cost canopy access equipment has made canopy study more viable as an option for scientific research. There are now a wide variety of access tools from which to choose, depending on the questions being addressed and the available budget (Lowman and Nadkarni, 1995). With the development of effective technological climbing methods such as the "canopy raft" and the canopy crane, and of ground-based methods such as insecticidal fogging, researchers now spend less time working on how to prudently work in the treetops and more time pondering the difficulties in recording meaningful canopy data, analyzing it, and interpreting the results.

A remarkable burgeoning of scientific interest in the canopy has occurred within the last decade. This is related to increasing concerns with such conservation issues as biodiversity, global atmospheric change, and management of tropical rain forests. The number of scientific publications on canopy structure has grown at a disproportionately rapid pace relative to the general field of biology (Nadkarni, 1994). Aspects of the canopy have been the focus of many recent symposia, scientific books, and popular articles and media.

Reviews of vascular epiphyte, hemiepiphyte, and parasite diversity have been compiled (Madison, 1977; Calder and Bernhardt, 1983; Kress, 1986; Benzing, 1990; Putz and Mooney, 1991; Williams-Linera and Lawton, 1995; Lowman and Nadkarni, 1995). The biodiversity of nonvascular plants and lichens has received less attention; only a single (but extensive) review of canopy cryptogams has been published (Rhoades, 1995). The biodiversity of canopy nonlichen fungi has not been well documented (D. Reynolds, pers. comm.).

For this review, we compiled the foregoing sources and searched the primary literature for additions and modifications. We also consulted on-line databases and communicated with numerous taxonomists and specialists to ensure that the information presented is current and accurate. To place canopy plant diversity into the context of biodiversity in a given study area, we compared inventories of epiphytes and nonepiphytes. The epiphyte quotient (sometimes called "epiphytic index") is defined as the percentage of epiphytes out of the total number of vascular plants in an area.

II. CATEGORIES OF CANOPY PLANTS

A. Vascular Epiphytes

Forest canopies support extensive flora that include over 24,000 species, or about 10% of all of the tracheophytes (Kress, 1986). Vascular epiphytes differ greatly in structure, function, and fidelity to their degree of dependence on canopy versus terrestrial habitats. Ecologists recognize their important roles in nutrient cycling and in providing arboreal and terrestrial animals with food, water, and nesting materials (Nadkarni, 1994). Ecophysiologists recognize the varied structures and mechanisms that protect vascular epiphytes from drought (Benzing, 1990).

Some of the characteristics for regular occurrence on bark and associated aerial substrates are obvious (e.g., holdfast roots and wind-dispersed propagules), but others are more subtle. In an extensive review of vascular epiphytism, Benzing (1990) outlined a variety of characteristics that are exhibited by vascular epiphytes (Table I).

Vascular epiphytes are mainly restricted to the low latitudes and within the tropics. They reach their greatest abundance and diversity at low to mid-montane elevations (Madison, 1977; Benzing, 1990). Ferns occur in higher latitudes along the margins of the Pacific, and a few hardy bromeliads and orchids occur in the mild north and south temperate zones (e.g., *Epidendrum rigidum, Polypodium polypodioides, Tillandsia usneoides*). The most extensively colonized temperate forests are those of southeastern Australia, New Zealand, and Chile, where a variety of vascular epiphytes grow in areas protected from frost by nearby warm ocean currents.

B. Nonvascular Epiphytes

In a recent review by Rhoades (1995), nonvascular (or cryptogamic) epiphytes were categorized into three groups: lichens, bryophytes, and free-living algae. Although the phylogeny and composition of the two plant groups considered here are very different—lichens are symbiotic fungi and algae, and bryophytes are plants—they occupy similar habitats and are often studied together. They have been the focus of little research in the upper canopy (relative to vascular epiphytes), except for a few studies in northwestern North America (Rhoades, 1995), the eastern deciduous forest of North America, and the boreal forest of Canada. Otherwise, most nonvascular epiphyte studies have been restricted to the lower trunks of trees and understory plants.

Bryophytes (phylum Bryophyta) are plants that lack true vascular tissues and organs. In canopy habitats they include the mosses (with about 10,000 species worldwide) and leafy liverworts (leafy hepatics, 7200 species). Thallose (strap- or fan-shaped) liverworts and hornworts are usually restricted to moist, lower trunks. (Rhoades, 1995).

Lichens are important components of canopy biodiversity and of ecosystem processes (e.g., nutrient cycling, providing food for wildlife). Lichens are not a

TABLE I

Characteristics of Vascular Epiphytes[a]

1. Reproduction	
A. Pollination	Exclusively zoophilous, flowers tend to be showy, pollinators highly mobile (Benzing, 1990, Chapter 5)
B. Breeding systems	Little studied, although many orchids appear to be allogamous
C. Population structure	Little studied
D. Seed dispersal	Most families endozoochorous, most species anemochorous (because of the dominance of Orchidaceae)
E. Life history	Almost all iteroparous, long-lived perennials
2. Vegetative	
A. Foliage	Usually evergreen, often succulent, and xeromorphic generally
B. Habit	Woody (wet forests) to herbaceous (wet and dry forests)
C. Shoot architecture	Various
D. Roots	Adventitious, specialized for holdfast, often reduced
E. Special features[b]	Impounding shoots (e.g., Bromeliaceae) and root masses (e.g., ferns), velamentous roots and absorptive foliar trichomes to prolong contact with precipitation and canopy washes, often lack capacity to grow in earth soil
3. Mineral nutrition	
A. Mycorrhizas	Possibly significant in Orchidaceae and Ericaceae, probably relatively unimportant elsewhere compared to terrestrial flora
B. Myrmecotrophy[b]	Nearly exclusive to epiphytes
C. Carnivory	Underrepresented in aboreal flora
D. Saprotrophy[b]	Phytotelm and trash-basket types
E. Special features[b]	Tolerance for low pH (wet forests), effective nutrient scavengers (dry forests), frequent reliance on organic substrates for nutrient ions
4. Photosynthesis/water balance	
A. Photosynthetic pathways	CAM overrepresented, no typical C_4 types, much interesting detail probably remains underdescribed
B. Water economy	Often very high
C. Moisture requirements	Various
D. Other[b]	Much flexibility, e.g., facultative CAM, CAM-C_4 intermediates

[a] Modified from Benzing (1990) with the permission of Cambridge University Press.

[b] These characteristics distinguish arboreal from terrestrial flora more than the others.

single taxonomically distinct category, but rather are symbiotic organisms, the association of a fungus (the mycobiont) and a photosynthetic partner (the photobiont), which are usually members of the Chlorophyta and Cyanobacteria. The degree of photobiont specificity varies among lichens. Generally, the mycobiont gives a lichen its overall form and provides the bulk of the biomass, outer protective layer, and a looser, inner layer that functions in physical absorption and storage of water and nutrients. The photobiont is usually restricted to the layer just below the protective covering of the mycobiont.

These cryptogams are poikilohydric, that is, they depend on an atmospheric supply of water and inorganic nutrients from precipitation, dew, or fog interception. In general, they absorb water rapidly and lack the water-resistant coverings or cuticles of vascular plants. Bryophyte growth forms have been described and discussed to understand their relationship to water use and conservation (Schofield, 1992; During, 1979). The gametophytes (vegetative bodies) of some species form tight cushions or spherical balls that expose a reduced surface area to retain water; others have a pendant, creeping habit that exposes them to maximal amounts of bark surface water. The sporophytes of many bryophytes are adapted to the periodically xeric nature of epiphytic habitats and can distribute their spores over very wide ranges (Gradstein *et al.*, 1989).

C. Hemiepiphytes

Hemiepiphytes have been defined as plants that have, at some point in their lives, an "umbilical" connection to the ground. Whether roots or stems, these connections buffer hemiepiphytes from problems of water and nutrient supply that are faced by obligate epiphytes (Williams-Linera and Lawton, 1995). Hemiepiphytes begin their life cycle either as epiphytes and eventually send roots and/or shoots to the ground (primary hemiepiphytes) or as terrestrially established seedlings that secondarily become epiphytic by severing all connections with the ground (secondary hemiepiphytes) (Kress, 1986).

Hemiepiphytes exhibit a tremendous variety in growth form, impact on their hosts, and degree of dependence on hosts. They range from being erect and treelike in form to species that grow in scandent, clambering heaps. Their impacts on hosts range from lethal (e.g., strangler figs) to benign (e.g., shrubby Ericaceae in tropical cloud forests) (Williams-Linera and Lawton, 1995).

D. Parasites

The mistletoes, which are woody shrubby parasites, are an ecologically distinctive group of canopy-dwelling plants. They have received a great deal of attention from botanists because of their ability to tap into the vascular system of their hosts, as well as from foresters, who have been concerned with reducing the damage they wreak through timber loss and mortality of desirable trees and shrubs (Calder and Bernhardt, 1983). Parasitic mistletoes tend to show a greater tendency for host specificity than do the epiphytes.

III. CANOPY PLANT TAXA DIVERSITY

A. Vascular Epiphytes

Although the global species richness of plants is probably in the region of 270,000, neither their exact number nor their global diversity pattern is known. It has been estimated that possibly as many 24,000 vascular plant species are epiphytes (Kress, 1986), so they constitute a major part of the global biodiversity in the forest canopy.

Vascular epiphytes account for 10% of the total vascular plant diversity. Most extant epiphytes are angiosperms, representing about 9% of all angiosperm species (Table II). Many vascular plant families (84) have adapted to life in the canopy, but relatively few taxa have radiated successfully. Within the angiosperms, approximately 31% of the monocots are epiphytic, whereas only 3% of the dicotyledons occupy the epiphytic niche. The Orchidaceae constitute approximately two-thirds of all epiphyte species (Kress, 1986). Other important monocotyledon families are Bromeliaceae and Araceae. The important canopy-dwelling dicotyledon families are Cactaceae, Ericaceae, Gesneriaceae, Melastomataceae, Moraceae, Piperaceae, and Rubiaceae (Table II; Kress, 1986). There are some large taxonomic groups of plants that contain no epiphytes or very few epiphyte species, for example, the Asteraceae, Leguminaceae, and Poaceae (Benzing, 1987). Less than 1% of the gymnosperms are known to be epiphytic. The Pteridophytes (ferns) are another important group of epiphytic plants, of which 29% are epiphytes (Kress, 1986).

TABLE II

Taxonomic Distribution of Vascular Epiphytes[a]

Major group	Taxonomic category	Number of taxa containing epiphytes in each category	Percentage of taxa containing epiphytes in each category
All vascular plants	Classes	6	75
	Orders	44	45
	Families	84	19
	Genera	876	7
	Species	23,456	10
Ferns and allies	Classes	2	67
	Orders	5	50
	Families	13	34
	Genera	92	39
	Species	2593	29
Gymnosperms	Classes	2	67
	Orders	2	33
	Families	2	13
	Genera	2	3
	Species	5	<1
Angiosperms (dicots)	Subclasses	6	100
	Orders	28	44
	Families	52	16
	Genera	262	3
	Species	4251	3
Angiosperms (monocots)	Subclasses	4	80
	Orders	9	47
	Families	17	26
	Genera	520	21
	Species	16,608	31

[a] Modified from Benzing (1990) with the permission of Cambridge University Press.

The epiphyte quotient (proportion of an entire flora that is epiphytic) varies widely both geographically and among forest types. This ratio has been measured directly in only a few study sites (Table III). Calculated epiphyte quotients based on published floristic studies in the Neotropics are known for Panama (12%), Peru (10%), Ecuador (22%), Costa Rica (26%), Venezuela (50%), and Florida (3%). In the Paleotropics, epiphyte quotients have been calculated from sites in Java (12%), West Malaysia (9%), Sri Lanka (4%), and Japan (0.5%). African forests seem to be much poorer in relative epi-

TABLE III

Epiphyte Quotients of South American Forests That Have Been Directly Measured by J. Nieder and His Colleagues

Study site	Elevation	Precipitation	Number of epiphyte species	Epiphyte quotient (area of reference)
Sehuencas, Bolivia	2100–2300 m	5000 mm/year	230 spp.	37% (0.1 ha)
Otonga, Ecuador	1700–2200 m	2500 mm/year	196 spp.	—
Rio Guajalito, Ecuador	1800–2200 m	2700 mm/year	166 spp.	28% (400 ha)
Carbonera, Venezuela	2100–2300 m	1500 mm/year	192 spp.	45% (360 ha)
Surumoni, Venezuela	100 m	2800 mm/year	53 spp. (crane plot only); 112 spp.	6 —

phyte species richness. In Ghana, a typical epiphyte quotient in forest plots is 8%; one direct measurement in central Africa (Rwanda and Zaire) was 3%.

It is generally regarded that the New World supports greater vascular plant diversity than the Old World. The number of vascular families containing at least one epiphyte species is very similar in the Paleotropics (43) and Neotropics (42). Within the Paleotropics, the representative families do not exhibit a homogeneous distribution. All 43 of the families occur in Australasia, but only 15 are found in Africa and Madagascar. Vascular epiphytes are most diverse in the Neotropics, and less so in tropical Asia and Africa. There is approximately a twofold increase in species diversity in the Neotropics compared with Australasia, and a sixfold increase compared with Africa (Madison, 1977; Gentry and Dodson, 1987). Vascular epiphytes exhibit their greatest diversity in the montane cloud forests of Latin America. The temperate regions support considerably fewer species than tropical areas. Likewise, the temperate regions generally support more vascular epiphytes than do boreal areas.

B. Nonvascular Epiphytes

In general, bryophytes account for 9–10% of the total species diversity of the plant kingdom. However, no one has calculated how many species of nonvascular plants are obligate epiphytes, as "the idea of an obligate epiphyte is a slippery concept" (D. Griffin, pers. comm.). Rhoades (1995) has ably summarized the results of regional floristic studies (Table IV).

The standard growth forms of lichens are arbitrary, but have often been used to describe functional groups in canopy habitats. "Foliose" refers to leaflike, "fruticose" refers to thalli without distinctive dorsoventral arrangements, and "crustose" refers to thalli firmly cemented to a substrate. According to the International Code of Botanical Nomenclature, lichen species are given the name of their mycobiont; photobiont names are subsidiary. Morphology (sexual structures, asexual structures, and vegetative surface characters) and thallus chemistry are important species characters. Only a few studies have focused on the worldwide biogeography of bryophytes (Schofield, 1992) or lichens (Rhoades, 1995). For many inventories, the crustose lichens have been lacking or incomplete, which is unfortunate as they are the dominant cryptogamic form in outer canopies. The proportion of lichens that grow arboreally is unknown.

The bryophytes are a very old group of plants, perhaps dating as far back in the fossil record as the Devonian period when the first land plants are known to have existed. The combination of a long history and small airborne dispores has allowed several bryophyte families to show wide geographic ranges. The tropical regions of Australia and Asia generally have more endemic genera of mosses, whereas in the Neotropics endemic liverwort genera are richer (Schofield, 1992). The cosmopolitan families are not restricted by latitude, but may show local altitudinal variation in some parts of their range. Representative moss families in the tropics include Bryaceae, Dicranaceae, Fissidentaceae, Funariaceae, and Hypnaceae. Pantropical moss families include Calymperaceae, Pterotyaceae, Racopilaceae, and Rhizogoniaceae. Important temperate families are Aulacomniaceae, Encalyptaceae, Grimmiaceae, and Polytrichaceae. Representative tropical liverwort families include Frullaniaceae, Lejeuneaceae, Lophocoleaceae, Plagiochilaceae, and Radulaceae. Species-rich liverwort fami-

TABLE IV

Species Richness of Epiphytic Cryptogams in Worldwide Forest Types[a]

Location/forest type	Latitude (°N)	Number of trees sampled	Mosses	Liverworts	Total bryophytes	Macrolichens
Guyana: dry evergreen *Eperua* spp.	5?	11	28	53	81	33
French Guyana; mixed lowland rain forest	5	4	43	61	104	21
Colombia; montane rain forest, 1500 m	5	4	22	36	58	49
Colombia; montane rain forest, 2550 m	5	4	33	102	135	51
Colombia; montane rain forest, 3510 m	5	4	19	63	82	37
Guyana; mixed lowland rain forest	7?	5	28	60	88	19
Oregon, United States; low, mixed coniferous forest	44	11	11	6	17	37
Wisconsin, United States; mixed conifers and hardwoods	46	Many	14	3	17	29
Montana, United States; old-growth *Abies*	48	5	4	1	5	34
Montana, United States; managed, second-growth *Abies*	48	5	1	0	1	37
Washington, United States; low-elevation fir forest on lava flow	49	Many	8	5	13	53
Sweden; deciduous forest	56	Many	78	17	95	—

[a] Modified from Rhoades (1995).

lies in temperate regions are the Marsupellaceae and Scapaniaceae. All the species of hornworts are in a single family, Anthocerotaceae, which is most diverse in tropical ecosystems (Schofield, 1992).

C. Hemiepiphytes

The phylogenetic distribution of hemiepiphytes suggests that this habit has evolved independently a number of times (Putz and Mooney, 1991; Williams-Linera and Lawton, 1995). Twenty-five families and 59 genera contain hemiepiphytes (Table V), with more than 820 species of primary hemiepiphytes and 650 species of seconary hemiepiphytes. These make up 1% of the total vascular plant species diversity, and 1% of the total canopy-dwelling vascular plant species. This is probably an underestimate, especially for woody hemiepiphytes ("lianas"), which are the most undercollected major canopy plant group. The stranglers most commonly occur in Moraceae and Clusiaceae, but are also found in Araliaceae, Rubiaceae, and Myrtaceae. The hemiepiphytic habit may have arisen from plants growing on rocks.

All of the hemiepiphytic monocotyledonous plants are secondary hemiepiphytes in the families Araceae and Cyclanthaceae. Secondary hemiepiphytes also occur in the dicotyledonous family Marcgraviaceae. Primary hemiepiphytes are represented by 20 families of dicotyledons. The majority of primary hemiepiphyte species are found in the families Araceae, Clusiaceae, and Moraceae. The Moraceae contain the most species of hemiepiphytes, with approximately 500 species in the genus *Ficus*. Primary hemiepiphytes (whose aerial roots eventually reach the ground) represent about 0.8% of all epiphytes with almost 2000 species (Gentry and Dodson, 1987).

As with vascular epiphytes, the contribution of hemiepiphytes to the diversity of the tropical forest canopy varies among forests. The percentage of trees colonized by hemiepiphytes has been reported for study sites in Venezuela (10% and 13%), Zimbabwe (13%), French Guiana (17%), and the Ivory Coast (21%). In neotropical lowland forests, stranglers and large hemiepiphytes can occur on 10–15% of the trees. Stranglers can occur in much higher densities in some neotropical palm savannas. A rich hemiepiphytic flora is typical of mountain forest and cloud forest sites, but wet lowland forest can also show high percentages of these species (25% in the case of La Selva Biological Station in lowland Costa Rica). In dry forests, hemiepiphytes are usually not present (Williams-Linera and Lawton, 1995).

Woody lianas are distinct features of tropical forests,

TABLE V

Families and Genera That Contain Hemiepiphytes[a]

SECONDARY HEMIEPIPHYTES

Monocotyledonae

1. Araceae
    - *Amydrium* Schott, 4/4 Malaysia
    - *Anthurium* Schott, 200/550 Neotropics
    - *Caladiopsis* Engl., 2/2 South America
    - *Epipremnum* Schott, 15/15 Indomalaya
    - *Monstera* Adans, 24/25 Neotropics
    - *Pedicellarum* Hotta, 1/1 Borneo
    - *Philodendron* Schott, 133/275 Neotropics
    - *Porphyrospatha* Engl., 3/3 Neotropics
    - *Pothos* L., 25/75 Indomalaya and Pacific
    - *Rhaphidophora* Hassk., 100/100 Indomalaya and Pacific
    - *Syngonium* Schott, 18/25 Neotropics
2. Cyclanthaceae
    - *Asplundia* Harling, 20/82 Neotropics
    - *Carludovica* Ruiz & Pav., 1/3 Central America
    - *Ludovia* Brongn., 2/2 South America
    - *Sphaeradenia* Harling, 7/38 Neotropics
    - *Thoracocarpus* Harling, 1/1 South America

Dicotyledonae

3. Marcgraviaceae
    - *Caracasia* Szyszyl., 2/2 Venezuela
    - *Marcgravia* L., 50/55 Neotropics
    - *Norantea* Aubl., 20/35 Neotropics
    - *Souroubea* Aubl., 20/25 Neotropics
    - *Ruyschia* Jacq., 2/10 Neotropics

PRIMARY HEMIEPIPHYTES

4. Araliaceae
    - *Didymopanax* Decne. & Planch, Neotropics
    - *Oreopanax* Decne. & Planch., Neotropics
    - *Pentapanax* Seem., 2/15 Java to Formosa
    - *Polyscias* J.R. & G. Forst, 5/80 Malaya to New Zealand
    - *Schefflera* J.R. & G. Forst, 60/200 Pantropics
    - *Sciadophyllum* P. Br., 5/30 South America and West Indies
    - *Tupidanthus* Hook.f. & Thoms., 1/1 Indomalaya
5. Bignoniaceae
    - *Schlegelia*[b]
6. *Burseraceae*
    - *Bursera*, 1 Costa Rica
7. Celastraceae
    - *Euonymus* L., 2/175 Himalayas
8. Clusiaceae
    - *Clusia* L., 85/145 Africa, Madagascar, Neotropics
    - *Clusiella* Planch. & Triana, 3/7 South America
    - *Havetiopsis* Planch. & Triana, 3/7 South America
    - *Odematopus* Planch. & Triana, 1/10 South America
    - *Quapoya* Aubl., 1/3 South America
    - *Renggeria* Meisn., 1/3 Brazil

PRIMARY HEMIEPIPHYTES, continued

9. Cunoniaceae
    - *Ackama* A. Cunn., 1/3 New Zealand
    - *Weinmannia* L., 3/170 New Zealand and Neotropics
10. Dulongiaceae
    - *Phyllonoma* Willd. ex Schult., 1/8 Neotropics
11. Ericaceae
    - *Cavendishia* (2 spp.), Neotropics
    - *Gonocalyx*, Neotropics
    - *Disterigma*, South America
    - *Sphyrospermum*, South America
12. Euphorbiaceae
    - *Schradera* (2 supp.)
13. Gesneriaceae
    - *Drymonia* (2 sp.), Central America
14. Griseliniaceae
    - *Griselinia* Forst.f., 3/6 New Zealand and Chile
15. Melastomataceae
    - *Blakea* P. Br., 60/70 Neotropics
    - *Topobea* Aubl., 20/50 Neotropics
16. Moraceae
    - *Coussapoa* Aubl., 20/45 Neotropics
    - *Ficus* L., 500/800 Pantropics
17. Myrsinaceae
    - *Grammadenia* Benth., 6/15 Neotropics
18. Myrtaceae
    - *Metrosideros* Banks ex Gaertn., 3/60 New Zealand
19. Potaliaceae
    - *Fagraea* Thunb., 20/35 Malaysia-Pacific
20. Rubiaceae
    - *Posoqueria* Aubl., 1/15 Neotropics
    - *Cosmibuena* Ruiz & Pav., Neotropics
21. Rutaceae
    - *Zanthoxylum*, Central America
22. Saxifragaceae
    - *Hydrangea*, Neotropics
23. Solanaceae
    - *Markea*,[b] Neotropics
24. Violaceae
    - *Melicitus*
25. Winteraceae
    - *Drimys*

---

[a] Modified from Williams-Linera and Lawton (1995). Numbers indicate the number of species/number of total species in the genus. After Madison (1977).

[b] From Putz and Mooney, 1991.

rarely or never occurring in temperate forests. Lianas account for approximately 10% of the tropical flora worldwide. They occur in greatest density in Madagascar and Africa, and less so in neotropical and Australian forests. At some sites in Madagascar, an average of 122 lianas with >2.5 cm dbh per 0.1 ha is reported, whereas lianas average only five individuals per 0.1 ha in northern temperate forests.

Lianas are more abundant in south temperate forests than in north temperate forests. The important families in north temperate forests are Anacardiaceae, Araliaceae, and Vitaceae. South temperate forests support more than twice as many families. Families like Bignoniaceae, Gesneriaceae, Gramineae, Lardizabalaceae, Saxifragaceae, and Vitaceae account for the majority of the climbing species (Putz and Mooney, 1991).

D. Parasites

About 1400 species of mistletoe occur in forests, woodlands, and shrublands on every continent except Antarctica, with most species in the tropics. Less than 1% of the total vascular plant species are mistletoes, and this group accounts for less than 1% of the total canopy-dwelling vascular plant species.

Mistletoes occur in two plant families. The Loranthaceae contain approximately 900 species in 65 genera, and the Viscaceae contain 400 species in 7 genera. The most species-rich genera in the Viscaceae are *Phoradendron* (170) and *Viscum* (100). The Viscaceae contain four genera restricted to the Old World, two genera that occur only in the New World, and one predominately New World genus is also widespread through Eurasia and Africa. In the Viscaceae, the New World genera *Dendrophthora* and *Phoradendron* contain about half of the 397 species of the family.

IV. DIVERSITY OF HABITATS OF CANOPY PLANTS

A. Canopy Microclimate

The values of light intensity and quality, temperature, wind, moisture content, and concentrations of various gases and aerosols are strongly modified by canopy structure in several ways. Canopy surfaces act as passive bodies for the absorption of wind energy, the dissipation of turbulence, and the sorption of heat and radiation. They also actively participate in exchanges of biologically important compounds, such as CO_2 and water vapor, which in turn may have an impact on regional, and even global, climate. Canopies also act as "filters" that remove small-scale turbulence, but allow large eddies to penetrate (Parker, 1995). Canopy structure therefore has a direct effect on the climate surrounding individual leaves, on the modification of microclimate through the layers of the forest, and on the large-scale environment of forest regions.

Canopy conditions are generally typified by more intense sunlight, greater extremes of relative humidity, higher water stress, and a smaller, more pulse-supplied pool of nutrients than on the forest floor. Sunlight attenuation can be as great as 98% between the tops of emergent trees and the levels reaching the forest floor. Rates of evaporation in the canopy have been recorded that are comparable to those occurring in open savannas. Relative humidity can range from nearly 100% at night to less than 30% during midday in the dry season. Differences in canopy versus forest floor wind speeds can also be extreme, especially in tropical cloud forests. In one Costa Rican ridge cloud forest, wind speeds within the canopy (10 m) were clocked at 11.3 m s^{-1}, while forest floor (2 m) speeds were only 4.0 m s^{-1} (Williams-Linera and Lawton, 1995).

B. Spatial Scales of Canopy Plant Diversity

The forest canopy is a three-dimensional subsystem of the forest itself. Canopy plants need relatively little space in order to develop a striking diversity. In an Ecuadorian montane forest, for example, 109 epiphyte species occurred on just 20 m^2 of branch surface, compared to only 67 terrestrial plant species on a 100-m^2 ground plot of elongated shape in the immediate vicinity of the phorophyte. The amazing concentration of epiphytes on single trees has often attracted the attention of naturalists. For example, 66 epiphyte species were found on one specimen of *Decussocarpus rospigliosii* in the Carbonera Forest in Venezuela.

The canopy offers its occupants a wide variation in water, light, and nutrient regimes compared to the understory and the forest floor, and this variety undoubtedly contributes to arboreal plant diversity. For example, soillike deposits and litter in the canopy function as a medium for canopy-dwelling plants that have well-developed root systems, such as vascular epiphytic shrubs. This material has a high organic content and is derived principally from leaf litter, feces, and other faunal remains. Also present is a small mineral component derived from fine particles carried there by wind, fog, and rain.

In tropical America, epiphytic bromeliads increase the volume of arboreal soil and litter by creating water-filled tanks ("phytotelmata") in which litter accumulates and soil forms. Although these arboreal epiphytes and their associated soils are patchily distributed, they are linked by climbing vegetation, by percolating rainwater, and probably by the movement of animals, and so they can be likened to a three-dimensional matrix of interconnected islands.

Within a forest, microsite differences exist at many different spatial scales: within a single branch; between branches at different heights of the tree; between trees of different architecture; and within stands of differing topography and aspects. Some studies have described variation in the distribution of canopy plants within single crowns. In a West African rain forest, for example, more than 75% of the orchid species grow on the inner branches, 48% in a middle zone, and only 4% in the outer canopy. Some research has linked differences in microsite water regimes and levels of sunlight input to differential distribution of certain canopy plants between outer and inner crowns; other arboreal species display a "generalist" distribution.

C. Host Tree Specificity

Studies of the mechanisms influencing host tree specificity are scant. It has been suggested that bark texture and pH influence patterns of colonization. For example, the orchid *Cymbidiella pardalina* obligately grows exclusively on the fern *Platycerium madagascarense*, whereas other epiphytes are found on numerous host tree species. This phenomenon warrants further investigation.

V. DIVERSITY OF GROWTH HABITS IN CANOPY PLANTS

Canopy-dwelling plants exhibit a great diversity of ecological adaptations, which is most likely a result of their diverse phylogenetic origins and the possibilities for adaptive specialization in the canopy habitat. The one unifying feature of these mechanically dependent plants is their occurrence in the canopy (Benzing, 1987). For vascular epiphytes, the different "habits" can be classified by several criteria, including degree of dependence (obligate vs. facultative), nutritional dependence (parasitic vs. commensalistic), degree of light demand (heliophiles vs. sciophytes), architecture (tank vs. atmospheric), substratum (e.g., ant-gardens, humiphiles), or carbon fixation pathway (Crassulacean acid metabolism [CAM] vs. C_3) (Benzing, 1990). Similarly, the great diversity of habits within the hemiepiphytic species ranges from stranglers, which ultimately become free-standing trees, to epiphytes that have only one root connecting with the ground.

A. Diversity of Physiology

The C_3 photosynthetic pathway is more typical of epiphytes inhabiting the canopies of cloud forests and cool, shaded, humid microsites. Forest canopies with more arid conditions favor CAM plants. The occurrence of a C_4 epiphyte has not been documented. Some families containing C_4 plants such as Asteraceae, Cyperaceae, Orchidaceae, and Poaceae also contain epiphytic species, but none is known to exhibit C_4 photosynthesis.

B. Diversity of Modes of Resource Acquisition and Retention

Nutrient acquisition in canopy plants occurs though many modes, including rainwater, bark and leaf leachate, nitrogen-fixing cyanobacteria, airborne particles, carton nests, crown humus, and decomposition of the host.

1. Humus Epiphytes

Ecologically, humiphily is the most common form of nutrient acquisition and supports the greatest diversity of epiphytes. An overwhelming majority of families containing at least one epiphytic species have at least one humiphile species, and most families contain only humiphile species (Benzing, 1987). Humus-rooted epiphytes include Ericaceae, Gesneriaceae, Melastomataceae, Piperaceae, and Rubiaceae, as well as the hemiepiphytes in Moraceae and Araliaceae. Many of the humus epiphytes are facultative. The obligate species exhibit a diversity of xeromorphic adaptations, including leaf succulence, a flattened pendulous growth form, and poorly developed root systems with strong mycorrhizal associations.

2. Tank and "Trash-Basket" Epiphytes

Some bromeliads, ferns, and orchids collect water, airborne particles, and leaf litter in the rosette created by overlapping fronds or leaves. Ferns in the genus *Asplenium*, for example, have roots that grow in the form of a trash basket to gain access to nutrients.

3. Ant-Associated Epiphytes

Myrmecophily is a common feature of vascular epiphytes and may be strictly a canopy phenomenon. Nest garden species include Araceae, Bromeliaceae, Cactaceae, Gesneriaceae, Marcgraviaceae, Orchidaceae, Piperaceae, and Rubiaceae. In Australia, Asclepiadaceae, Melastomataceae, and Rubiaceae occur in less-studied ant-garden symbioses. Asclepiadaceae, Bromeliaceae, Melastomataceae, Orchidaceae, Polypodiaceae, and Rubiaceae contain specialized ant-garden species. Rubiaceae appear to be the most specialized ant-garden species. Some ferns and orchids have hollow rhizomes (Polypodiaceae), hollow tubers (*Solanopteris* spp.), and hollow pseudobulbs (*Schomburgkia* and *Laelia*) that provide domatia for ants.

4. Bark Epiphytes

To inhabit the bark substrate, epiphytic plants must cope with very low levels of water and nutrient availability. Therefore, many of the bark epiphytes are obligate epiphytes, including many specialized orchids.

5. Atmospheric Epiphytes

Some bromeliads (e.g., *Tillandsia*) have special hairs (trichomes) that allow them to absorb water from the atmosphere over the entire surface of their leaves.

VI. CANOPY PLANT BIODIVERSITY AND CONSERVATION BIOLOGY

A. General Considerations of Canopy Plant Conservation

Because of their small size, high degree of endemism, and frequent microsite specificity, epiphytes may be more vulnerable to human-induced disturbance than terrestrial plants. Methods to conserve existing epiphyte populations and floras have been discussed (e.g., Lowman and Nadkarni, 1995). Studies have shown the value of older trees in forests as habitats for certain sensitive species.

B. Effects of Forest Fragmentation and Habitat Conversion

The effects of forest fragmentation and habitat conversion on canopy plant diversity are poorly documented, especially in the tropics. It is generally accepted by researchers that secondary bryophyte and lichen communities are very different from those in primary forests (Gradstein *et al.*, 1989). Most studies indicate a decrease in species richness between secondary habitats and primary forests, and even disturbances at small spatial scales (within a branch) are reported to result in a decrease in diversity.

Shade epiphytes growing in the understory are more affected by habitat conversion than the sun epiphytes of the canopy, but not all sun epiphytes are able to recolonize following disturbance. The available data from investigations of the regeneration rates of temperate and subtropical canopy plants indicate that many species are slow to recover. The rates for bryophytes range from 25 years in Australia to 80–100 years in California. In Britain, it is estimated that lichens may require 500 years to successfully regenerate.

C. Effects of Global Environmental Change

Water stress is a major limiting factor for plants inhabiting the crowns of trees. A rise in global temperatures may have an impact locally on the relative humidity of some forest canopies. Preliminary experimental work along an altitudinal gradient indicates that the species composition of canopy plant communities may be altered by such changes in temperature and humidity.

D. Vulnerability of Canopy Plants to Extinction and Invasion

There are no records of a specific canopy plant extinction in modern times. However, numerous endemic species are endangered or threatened by habitat conversion (Gradstein *et al.*, 1989). A great deal of information indicating that lichens are very susceptible to air pollution and metal ion deposition has accumulated over the past several decades (Rhoades, 1995). Gradstein *et al.* (1989) suggest that relatively small reserves containing a diversity of life zones should suffice to conserve cryptogam biodiversity if the reserve is large enough to maintain a viable population of host trees. However, they warn that these recommendations are based on very preliminary data and more inventory data and taxonomic work are needed to better define species ranges and to determine which species are locally rare or endemic. Mistletoes are generally very susceptible to environmental changes (Calder and Bernhardt, 1983).

VII. AREAS FOR FURTHER STUDY

A. General Considerations

Habitat loss and climate change are growing threats to plant communities. Arboreal plants provide many opportunities and challenges for biologists from many disciplines, and because there plants have no access or sporadic access to terrestrial soil, they make excellent experimental subjects to study physiology and stress. Canopy plants warrant attention for the roles they play in forest dynamics, which affect biodiversity, productivity, and nutrient cycling. A list of research questions was created for vascular epiphytes (Table VI); these questions can also be related to the study of other types of canopy plants.

B. Monographs and Inventories

There is a pressing need for extensive and intensive work on plants that live in the canopy. However, the lack of resident tropical taxonomists is a serious concern. There are many more taxonomists in the more developed countries where the resources and infrastructure exist to train students, but there are relatively few specialists in less developed countries where many of the biological resources exist.

However, efforts to create monographs and inventories of canopy-dwelling plants have been increasing. For example, botanists at the Missouri Botanical Gardens and their collaborators have compiled inventories of regional floras. To date, Peru, Panama, Venezuela, and the Guianas have received a great deal of attention, and the study of other floras (e.g., Nicaragua and China) is planned. To fully represent regional biodiversity, it is crucial that botanists collect plants in the canopy. Likewise, to fully understand global biodiversity, the generally undercollected groups must be collected in the canopy (e.g., lianas and cryptogams). In terms of cryptogams, work needs to continue on broad regional inventories of all tropical species and of crustose lichens worldwide. The bryoflora of Australia is particularly poorly known.

C. Herbaria and Databases

Certain herbaria have significant canopy plant collections (Madison, 1977; Kress, 1986). Herbarium studies have been conducted at the Harvard University Herbaria, the Marie Selby Botanical Gardens, the Huntington Botanical Gardens, the Herbario Nacional Colombiano, the State University of Utrecht herbarium, and the University of Florida.

Several major botanical gardens have produced useful databases. For example, The Missouri Botanical Gardens has developed an on-line database (TROPICOS) for the floras of several Neotropical countries. Such databases can be of great use to systematic biologists and conservationists by providing the most up-to-date information available. With the increasing attention being paid to canopy plant ecology (e.g., Lowman and Nadkarni, 1995), perhaps it is time to initiate a canopy plant biodiversity database.

D. Experimental Fieldwork

Experimental field studies to investigate the potential effects of forest harvesting on plant community composition and species richness should be conducted, especially in tropical regions, so as to include biodiversity objectives in forest management practices. Humus epiphyte communities growing in bryophyte mats, for example, are ideal for experimental fieldwork because entire moss mat communities are easily transplanted with minimal disturbance to the rooting medium. Transplanting epiphytes along an altitudinal gradient is useful in helping to predict the effects of environmental change. The experimental removal of bryophyte mats is also useful in monitoring epiphyte succession and recolonization.

E. Ethnobotany of Canopy Plants

Ethnobotanical knowledge and usage of canopy plants is widespread in cultures around the world (Lowman and Nadkarni, 1995). However, the pharmaceutical potential of canopy plants has only begun to be investigated. In particular, the bryophytes and lichens have not been rigorously explored in this regard. The potential importance of canopy plants for human use may spark resources needed to learn more about both applied and basic aspects of these diverse organisms.

See Also the Following Articles

FOREST CANOPIES, ANIMAL DIVERSITY • PLANT BIODIVERSITY, OVERVIEW • TROPICAL ECOSYSTEMS

Acknowledgments

This summary of canopy plant diversity was supported by the National Science Foundation. We thank the Monteverde Cloud Forest Reserve and the Tropical Science Center in San José, Costa Rica for main-

TABLE VI

Research Questions and Opportunities for Canopy Plants[a]

Subject	Obvious	Questions remaining
1. Fidelity to canopy versus other substrates	Occurrence on trees ranges from accidental to obligate.	What factors differentiate canopy from terrestrial substrates for the obligate epiphyte? How has specialization for arboreal life compromised capacity to survive on the ground?
2. Requirements for specific types of arboreal substrates	Specific epiphytes typically colonize only subsets of the many types of substrates present in occupied tree crowns.	What plant characteristics determine microsite requirements for twig, bark, humus, ant-nest garden, etc., epiphytes?
3. Plant adjustments to the often transitory and relatively unpredictable supplies of moisture in forest canopies	Broadly occurring accommodations to drought (e.g., CAM, xeromorphy) are particularly well developed among the epiphytes.	What is the nature of the moisture supply in forest canopies and how are mechanisms such as photosynthetic pathways, osmotic balance, and stomatal behavior fine-tuned to reduce risk and maximize effective use of available moisture?
4. Plant adjustments to the absence of mineral soil	A variety of organic substrates, including the products of mutualistic biota, serve in lieu of earth soil as sources of nutritive ions.	How is impounded litter processed for phytotelm epiphytes? How substantially do ant mutualists contribute to the nutrient budgets of associated epiphytes? How are the more oligotrophic epiphytes (e.g., atmospheric bromeliads) equipped to scavenge scarce ions and use them economically?
5. Impacts of arboreal ants	Some epiphytes require ants for dispersal and to provide rooting media.	How much arboreal flora beyond the obvious ant-nest garden and myrmecotrophic species are dependent on ants for dispersal, substrates, and defense?
6. Epiphytic vegetation as a resource for canopy fauna	Much arboreal fauna, particularly invertebrates, use epiphytes as resources.	What is the full extent of this dependence and what are the broader consequences of these dependencies for the forest community?
7. Epiphyte involvements in nutrient cycles	Nutritional piracy exists. Epiphyte biomass sometimes contains much of the nutrient capital present in a forest ecosystem.	To what degree and under what conditions does the presence of an epiphyte load have an impact on the nutritional status of a phorophyte?
8. Impacts on community productivity and patterns of resource use	Resources present in epiphyte biomass (e.g., N and P) at least sometimes yield photosynthetic returns at different rates than those of supporting soil-rooted vegetation.	How does the presence of substantial epiphyte biomass affect aggregate forest productivity and help determine overall resource-use efficiency?
9. Conservation	Because many epiphytes occupy narrow ranges (especially orchids), often in regions of rapid development, endangered status is correspondingly common.	What conservation strategies are likely to preserve the greatest diversity of epiphytes?
10. Indicators of habitat quality and global change	Some epiphytes possess characteristics that impart extraordinary utility as air quality monitors.	How can epiphytic vegetation be more effectively used to monitor changing conditions in the troposphere?
11. Succession	Presumed seral stages identified.	Do species displace one another on bark? If so, by what mechanisms?
12. Community organization	Species often co-occur in predictable assemblages, but often distribution and spacing among individuals are random.	Are the factors responsible for the distributions and combinations of species on bark primarily density dependent or density independent?

[a] Modified from Benzing (1990).

taining and protecting the Research Area. Funds were provided by Evergreen State College Faculty Sponsored Research, the National Science Foundation [Ecology Program (BSR 86-14935 and 89-18006), Long-Term Studies Program (DEB 96-15341), Database Activities Program (BIR 96-30316)], the Whitehall Foundation, the National Geographic Society, and the Evergreen Foundation. Many plant taxonomists provided information on epiphyte floristics. We thank A. S. Taylor (Zamiaceae), T. Croat (Araceae), R. Faden (Commelinaceae), H. Luther and B. Holst (Bromeliaceae), H. Kenedy (Marantaceae), L. Thein (Winteraceae), P. Berry and P. Hoch (Onagraceae), P. Lowry and D. Frodin (Araliaceae), M. Endress (Apocynaceae), and W. J. Kress (for many families).

Bibliography

Benzing, D. H. (1987). Vascular epiphytism: Taxonomic participation and adaptive diversity. *Ann. Missouri Botanical Gardens* 74, 183–204.

Benzing, D. H. (1990). *Vascular Epiphytes.* Cambridge University Press, Cambridge, United Kingdom.

Calder, M., and P. Bernhardt (eds.). (1983). *The Biology of Mistletoes.* Academic Press, Sydney, Australia.

During, H. J. (1979). Life strategies of bryophytes: A preliminary review. *Lindbergia* 5, 2–18.

Erwin, T. L. (1988). The tropical forest canopy: The heart of biotic diversity. In *Biodiversity* (E. O. Wilson, ed.), pp. 123–129. Stanford University Press, Stanford, California.

Gentry, A. H., and C. H. Dodson. (1987). Diversity and biogeography of neotropical vascular epiphytes. *Ann. Missouri Botanical Gardens* 74, 205–233.

Gradstein, S. R., G. B. A. van Reenen, and D. Griffin. (1989). Species richness and origin of the bryophyte flora of the Colombian Andes. *Acta Botanica Neerlandia* 38, 439–448.

Kantvilas, G., and P. R. Michin. (1989). An analysis of epiphytic lichen communities in Tasmanian cool temperate rain forest. *Vegetatio* 84, 99–112.

Kelley, D. L. (1985). Epiphytes and climbers of a Jamaican rain forest: Vertical distribution, life forms, and life histories. *J. Biogeogr.* 12, 223–241.

Kress, W. J. (1986). The systematic distribution of vascular epiphytes: An update. *Selbyana* 9, 2–22.

Lowman, M. D., and N. M. Nadkarni. (1995). *Forest Canopies.* Academic Press, San Diego, California.

Madison, M. (1977). Vascular epiphytes: Their systematic occurrence and salient features. *Selbyana* 2, 1–13.

Nadkarni, N. M. (1994). Diversity of species and interactions in the upper tree canopy of forest ecosystems. *Amer. Zool.* 34, 70–78.

Parker, G. G. (1995). Structure and microclimate of forest canopies. In *Forest Canopies* (M. D. Lowman and N. M. Nadkarni, eds.). Academic Press, San Diego, California.

Putz, F. E., and H. A. Mooney. (1991). *The Biology of Vines.* Cambridge University Press, Cambridge, United Kingdom.

Rhoades, F. M. (1995). Nonvascular epiphytes in forest canopies: Worldwide distribution, abundance, and ecological roles. In *Forest Canopies* (M. D. Lowman and N. M. Nadkarni, eds.), pp. 353–408. Academic Press, San Diego, California.

Schofield, W. B. (1992). Bryophyte distribution patterns. In *Bryophytes and Lichens in a Changing Environment* (J. W. Bates and A. M. Farmer, eds.). Oxford University Press, Oxford, United Kingdom.

Williams-Linera, G., and R. O. Lawton. (1995). The ecology of hemiepiphytes in forest canopies. In *Forest Canopies* (M. D. Lowman and N. M. Nadkarni, eds.), pp. 255–283. Academic Press, San Diego, California.

FOREST ECOLOGY

Timothy J. Fahey
Cornell University

GLOSSARY

disturbance	An event that results in a radical change in environment, usually as a consequence of death of the dominant plants.
food web	A set of species that live together and a specification of which species "eat" which other species.
forest stand	A relatively homogeneous forest landscape unit that can be distinguished from neighboring units by forest age or composition.
forest structure	The arrangement of all the parts of the forest stand—stems, branches, leaves, roots, and so on.
keystone species	A species that performs some crucial activity in the ecosystem and for which there is little or no redundancy.
succession	Sequential change in the relative dominance of various species in the biotic community at a location.

A FOREST IS AN ecological system (or biotic community) dominated by trees. Ecology is the study of ecological systems and the mutual interactions between organisms and the environment. The primary objective of forest ecology is to understand what controls the patterns of distribution and abundance of different organisms in forests of the world. The relevance of forest ecology to understanding biodiversity lies in the observations that forests dominate the natural landscape over much of the world and that forests harbor a large proportion of the world's species. Moreover, because forests are valuable to humans for the products and services that they provide—especially wood—many forest ecosystems are extensively manipulated and modified by human societies, often disrupting the natural ecological patterns. In addition, many forests occur where climate and soil are suitable for intensive agricultural production and conversion of forests to farms results in radical changes in the composition of the biota. Hence, detailed knowledge of forest ecology is necessary to predict and ameliorate the effects of forest utilization and conversion on biodiversity.

I. INTRODUCTION

Forests occur naturally wherever environmental factors permit the upright woody growth form of trees to be a successful strategy. The primary advantage that trees realize in such settings is the ability to gain a competitive edge over any smaller neighbors in gaining access to solar radiation for photosynthesis. Thus, forests dominate the natural landscape in all regions except where the upright woody growth form is not possible or where other mitigating environmental factors eliminate any competitive advantage of being tall. For example, in the Great Plains of North America grass fires were usually too frequent to allow trees to become established and outcompete the grasses for sunlight.

Ecologists have classified the forests and other vegetation of the world into biological regions, or biomes, on the basis of the physiognomy (outward form) and species composition of the dominant plants. The biome classification represents ecological patterns at a very large scale. However, the patterns that ecologists endeavor to explain occur at a wide range of spatial scales, from global to local. Global scale maps of the world's biomes have been constructed to illustrate the broad distributional patterns of the biota. Ecological patterns at smaller scales are represented by subdividing the various biomes to provide more detailed classifications and maps. In essence, the forest ecologist regards the global ecosystem as consisting of a nested series of ecological associations, at smaller and smaller spatial scales, down to the level of the relatively uniform forest stand.

The focus of ecological classification systems on the dominant plants is not just a convenience; the dominant plants in terrestrial ecosystems (e.g., the trees in the forest) play a crucial role in setting the stage for all the other organisms because they convert solar energy into food for the food web and they form the three-dimensional structure that constitutes the habitat of most associated biota. Hence, it is of primary importance in studying the ecology of forests to understand what controls the distribution and abundance of trees. However, this is not to say that the other organisms are not important to the functioning of the ecosystem; in fact, many of the less prominent organisms play important roles in regulating the distribution and abundance of the trees.

Underlying these ecological classification systems and maps at the broad scale is the principle that climate is the environmental feature that exerts primary control over the distribution of organisms on the earth's surface. This principle was realized long ago by European geographers who observed during their explorations the coincidence between the broad patterns of climate and vegetation physiognomy on earth. The key components of climate, temperature and precipitation, exert this control through their effects on the growth of plants. Different plant traits, expressed in part as whole plant physiognomy, prove to be most suitable for growth and survival under different combinations of temperature and precipitation. Hence, climatic patterns largely coincide at the broad scale with the patterns of distribution of forest biomes. Expressed most simply, the effect of climate on vegetation distribution can be plotted in terms of mean annual temperature and precipitation (Fig. 1); forest vegetation is restricted to relatively moist climates. However, annual averages cannot account for the important effects of climatic seasonality—variations in temperature and precipitation through the year—in determining biome distributions. More complicated systems of expressing the influence of climate account for the important effects of seasonal drought and of subfreezing temperatures on plant growth and activity.

The principle that climate exerts primary control over the distribution of organisms fails to account for the observation that some equivalent climates support vegetation with differing composition and physiognomy. For example, the temperate rain forests of the Pacific Northwest of North America are dominated by evergreen needleleaf trees whereas in some equivalent climates of the southern hemisphere (New Zealand and South America) evergreen broadleaf forests may dominate. These differences reflect the fact that the flora and fauna that have developed by biological evolution in the various regions of the world are quite distinct because they have been geographically isolated from one another. Several biogeographic zones have been identified by botanists and zoologists, and these zones reflect geologic history, especially continental drift. For exam-

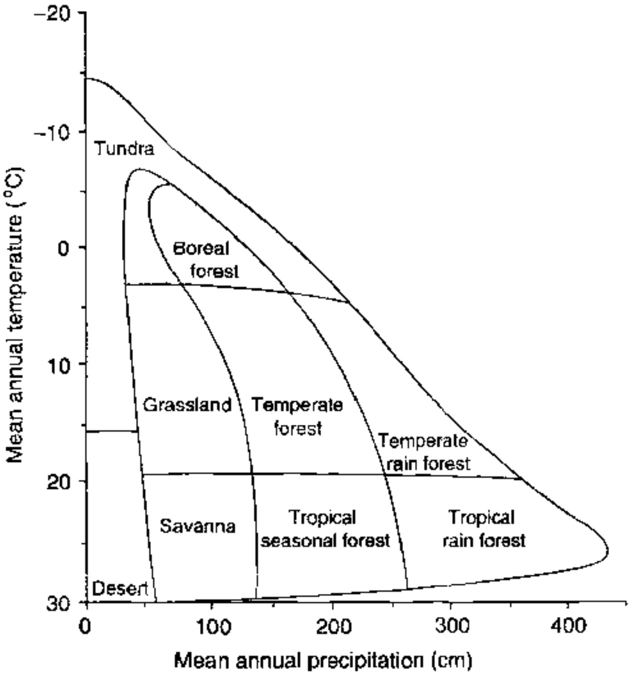


FIGURE 1 The affects of mean annual precipitation and temperature on vegetation. (Modified from Whittaker, 1975, with permission.)

ple, plant geographers typically recognize four plant domains whose floras are distinct from one another: (a) paleotropical (Old World Tropics), (b) neotropical (New World Tropics), (c) north temperate, and (d) south temperate. To expand on our earlier principle, then, climate together with the available flora and fauna exert primary control over the distribution and abundance of organisms on earth. This principle points toward the likelihood that human effects on climate and the introduction of exotic species across the globe will have profound consequences for ecology and biodiversity.

At a more localized scale, species distribution and abundance are also affected by secondary environmental factors, topography and soils. Topography exerts its influence in part by locally modifying climate; for example, in the northern hemisphere south-facing slopes are warmer and drier than north-facing slopes. Also, topography influences the environment through the action of gravity moving matter downhill (e.g., water, soil particles). Soil supplies the essential environmental resources, water and mineral nutrients, for tree growth. Soil properties reflect the combined effects of geologic, climatic, and biologic forces, emphasizing the complex web of cause and effect that underlies ecological patterns and processes.

Geologists have divided the earth's landscape into physiographic provinces that reflect the effects of geological processes on the earth's surface features. Within each physiographic province, relatively orderly and recurring patterns of topography and soils are observed that differ fundamentally from those of neighboring provinces. For example, in the coastal plain province of the eastern United States, sandy sediments have emerged from the receding ocean to leave a gently rolling landscape of porous, sandy soils. Forest patterns reflect subtle variations in topography and drainage. In contrast, directly to the west, the ridge-and-valley province is marked by recurring combinations of bold narrow ridges of resistant rocks and intervening valleys underlain by softer substrates like limestone and shale. The forest patterns in this province directly mirror the striking gradients in environment that result from the combination of topography and soils determined by the bedrock geology of the province. An understanding of the ecology of these two regions begins with the recognition of the underlying differences in geological forces that have shaped their physiography.

One additional factor that strongly regulates the composition of the biotic community is the legacy of disturbance events. Disturbance is a natural phenomenon in all ecosystems and is defined as any event that results in a change in environmental conditions and resource availability, usually as a consequence of death of the dominant plants. Among the most prominent natural disturbance agents in forest ecosystems are fire, windstorms, and irruptions of pests and pathogens. When disturbances occur at a spatial scale that is much larger than the area occupied by individual dominant plants (i.e., large-scale disturbances), they so profoundly alter the environment that the suite of plants that subsequently colonizes the disturbed site may be quite different from the original community. These large-scale disturbances initiate the ecological process of succession—that is, successive changes in the composition of the biotic community occur as a result of progressive changes in environmental factors. Hence, the actual composition of the biotic community at any particular time and place depends not only on climate, flora and fauna, topography and soils, but also on the time interval since the last large-scale disturbance that initiated successional change—as well as the nature and intensity of that disturbance event. That disturbance has played a major role in shaping ecological patterns and processes is evident from the many traits of the flora and fauna that reflect the selective force associated with disturbance.

The natural disturbance regime that characterizes any particular ecosystem depends on exogenous factors that act as disturbance agents (e.g., the combination of drought, lightning, and wind that favors the occurrence of fire; or frequent exposure to hurricanes or tornadoes in certain geographic regions) as well as endogenous factors, such as the traits of the plants themselves, that influence the frequency or intensity of the disturbance. For example, pine forests are more prone to fire disturbance than deciduous broadleaf forests in part because the fuels produced by pine trees are more flammable. And forests on sandy soils are more prone to fire because coarse soils dry out more rapidly than fine-textured soils. Thus, the influences of disturbances, environment, and biota may be mutually reinforcing in shaping ecological patterns and processes.

The prevalence of disturbance and succession in natural forest ecosystems complicates the task of classifying and mapping forest distributions, particularly at smaller scales, because forest composition is continually changing. Recognizing this problem, ecologists have defined the climax forest association as the assemblage of species that would persist under any particular combination of environmental factors in the absence of large-scale disturbance. In many regions, however, the recurrence of large-scale disturbances is naturally so frequent that climax forest associations rarely develop, and the practical value of such a strict climax concept is somewhat limited. However, in some regions where large-scale disturbances are infrequent, particularly under humid climates (where fires are rare) and areas not often exposed to intensive windstorms (i.e., the humid tropics and some temperate areas), climax forest associations probably were common prior to the advent of anthropogenic influences. Natural disturbance regimes in these forests consisted of small-scale events resulting from the death of individuals or small groups of trees in a patchwork mosaic. Although the composition and structure of various patches would differ, at a larger scale the average composition of the forest ecosystem would remain relatively steady (i.e., it would exhibit a shifting-mosaic steady-state). These observations illustrate the importance of the scale of observation to our understanding of ecological patterns and processes.

II. BIODIVERSITY IN FOREST ECOSYSTEMS

A. General Patterns

The number of species observed in any particular forest varies markedly across the earth's forest biomes. Biodiversity can be conceived as consisting of three distinct elements, termed the gamma, alpha, and beta diversity. The total diversity in a large area, the gamma diversity, can be partitioned into two components, the local (alpha) diversity in a single habitat or forest stand, and the turnover of species between stands, the beta diversity. High gamma diversity could be associated either with high alpha diversity, high beta diversity, or both. Ideally, considerations of the patterns of biodiversity across forest regions need to account for the contributions of these different elements of biodiversity. Unfortunately, studies of forest biodiversity patterns generally have not provided samples of sufficient detail to resolve these elements, and more systematic approaches are needed. Nevertheless, many valuable insights into the patterns and causes of biodiversity variation across forest regions have emerged from research to date.

The most striking pattern of forest biodiversity is the latitudinal gradient: much higher biodiversity is observed both locally and regionally in tropical than in temperate forests. Because observations are most comprehensive for tree species, the following description focuses on these taxa.

The general relationship between latitude and the alpha diversity of trees is illustrated by a plot of the number of species in 0.1 ha samples taken from around the world (Fig. 2). At one extreme are lowland tropical rain forests in the upper Amazon basin where on average every second tree in a sample belongs to a different species; at the other extreme, a variety of natural, monospecific forests are common in many temperate and boreal regions. Within the tropical region considerable variation in alpha diversity has been observed (Fig. 2). The two most prominent features correlated with this variation are annual precipitation and biogeographic province or region. For example, the average alpha diversity (number of tree species in 0.1 ha samples) of neotropical lowland wet and moist forests is 152 species, whereas for seasonally dry forests the average is 65. And the alpha diversity of trees in Africa is generally much lower than for eastern Asia and America (Fig. 3). The extremely high alpha diversity of lowland tropical rain forest is not paired with equally high beta diversity. For example, the 307 tree species identified on a single 1-ha plot in the Ecuadorian Amazon constituted about 16% of the total tree flora (trees ≥ 5 cm diameter) of Amazonian Ecuador and a single 50-ha plot at Pasoh, Malaysia contained 830 species, 20 to 30% of the total tree flora of this country.

By comparison, tree species diversity in the temperate latitudes is much lower: the enormous area of temperate zone forests harbors only 1166 tree species, not much more than the 50 ha plot at Pasoh! As for the tropics, however, striking geographic differences in the diversity of tree taxa are observed in the temperate zone. The highest gamma diversity is in east-central Asia,

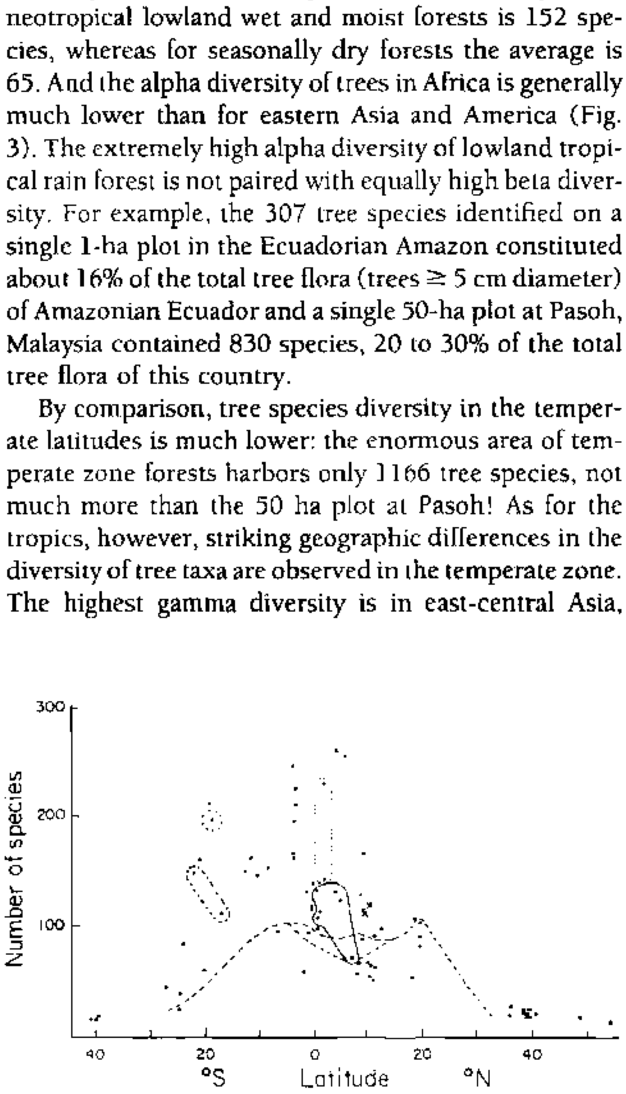


FIGURE 2 Species richness of 0.1 ha samples of lowland (<1000 m) forest as a function of latitude. Dashed line separates dry forest (bottom) from moist and wet forest (top) with intermediate sites (moist forest physiognomy despite relatively strong dry season) indicated by alternate lines. X, the anomalous Coloso and Loma de los Colorados sites in northern Colombia. Reprinted, with permission, from Gentry (1995).

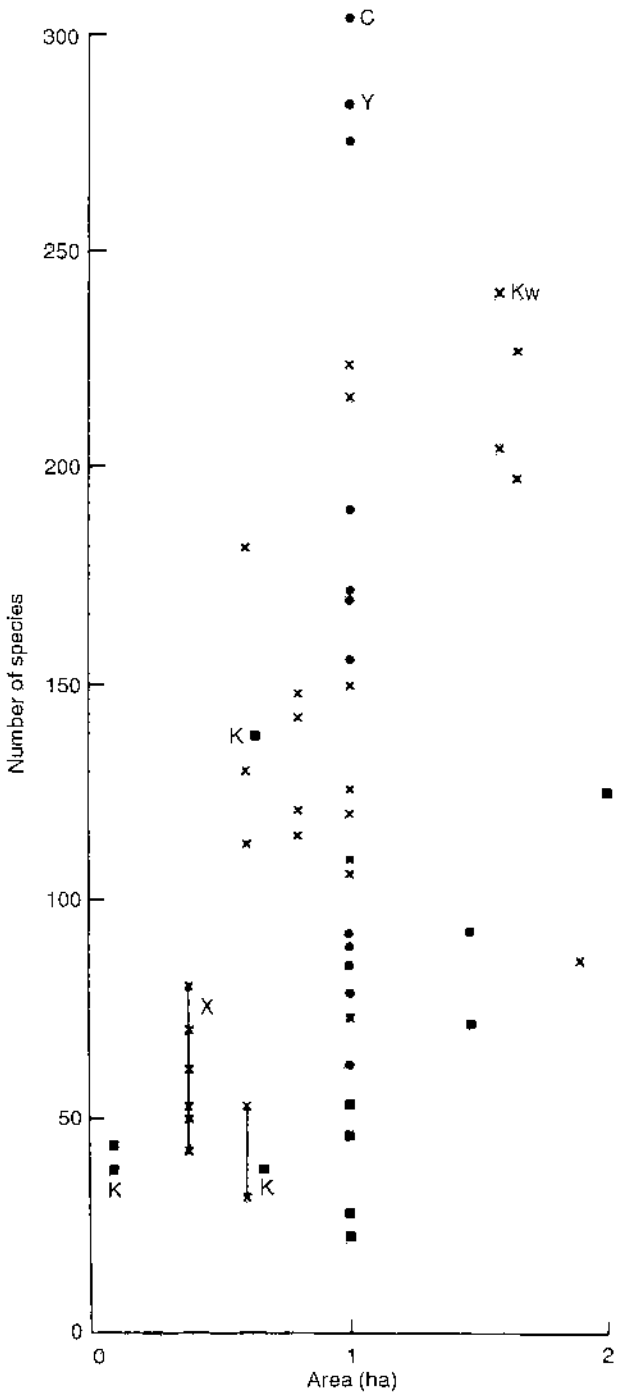


FIGURE 3 Number of species shown among trees of at least 0.1 m in diameter on small plots in tropical lowland rain forest. ●, America; x, Eastern tropics; ■, Africa. Lines connect sample plots that lie close together. Reprinted, with permission, from Whitmore (1998).

with intermediate levels in eastern North America and lowest in Europe (Table I).

A comprehensive explanation of these global patterns in tree species diversity remains elusive, but a growing consensus on the role of historical or biogeographic factors and local physical habitat factors is emerging. The overall lower diversity of temperate than tropical regions probably is explained in part by the physiological constraints on colonization imposed by the need to tolerate subfreezing conditions and by the smaller contiguous area in temperate regions. Regional differences between biogeographic provinces in the temperate zone appear to owe in large part to higher extinction rates in Europe and North America during the Pleistocene glacial epochs as well as greater access of the east Asian region to dispersal routes from the tropics. Similarly, the combination of geographic area and climatic differences probably explains the contrasts in diversity between tropical Africa versus America and east Asia and between wet and dry forests. As noted by Latham and Ricklefs (1993), "further resolution of the causes of diversity patterns will require new paleontological, biogeographical and taxonomic data and synthesis." (p. 310)

B. Forest Structure and Pattern and Disturbance Regimes

The local biodiversity in particular forests depends on the complex suite of factors that characterize the habitats of individual species. These factors include such components as the species composition, phenological timing, structural complexity, and horizontal patterning of the vegetation, which in turn depend on environment and the legacy of disturbances. The dominant plants (i.e., trees in forests) play a pivotal role in defining the habitats for associated organisms by providing food and shelter and by regulating the local microenvironment. For some organisms (e.g., many insects) host-specific interactions with particular dominant plants result in strong correlations in their distributions, so that forest composition is the key factor influencing the composition of associated species. For other organisms, the structural and horizontal patterning of the dominant vegetation may be more important than composition alone.

The importance of the three-dimensional spatial arrangement of the branches and leaves of the plants in defining animal habitats was realized by pioneers in the study of ecology and evolution. MacArthur and MacArthur (1961) demonstrated that the diversity of birds in forests of the eastern United States could be

TABLE I

Summary by Taxonomic Level and Region of Moist Temperate Forest Trees in the Northern Hemisphere

Taxonomic level	Number of tree taxa characteristic of moist temperate forests in:				
	Northern, central, and eastern Europe	East-central Asia	Pacific slope of North America	Eastern North America	Northern Hemisphere (total)
Subclasses	5	9	6	9	10
Orders	16	37	14	26	39
Families	21	67	19	46	74
Genera	43	177	37	90	213
Species	124	729	68	253	1,166
Families excluding those of predominantly tropical distribution (% of total)	18 (86%)	37 (55%)	18 (95%)	29 (63%)	41 (55%)
Genera excluding those of predominantly tropical distribution (% of total)	41 (95%)	121 (68%)	35 (95%)	77 (86%)	149 (70%)
Species exclusive of predominantly tropical genera (% of total)	122 (98%)	570 (78%)	66 (97%)	236 (93%)	987 (85%)

Data from Latham and Ricklefs, 1993.

predicted by the structural complexity of the vegetation (Fig. 4): in habitats with high foliage-height diversity (FHD; defined by the formula $\mathrm{FHD} = \%_i\, p_i \ln p_i$, where p_i is the proportion of total foliage area in the i^{th} layer), bird diversity was much higher than in habitats with low FHD. Moreover, this relationship largely transcended differences in plant species diversity. Similarly, for ground-dwelling organisms, the complexity of habitat at the soil surface—including vegetation cover, litter, rocks, fallen logs, and moisture—is strongly correlated with local biodiversity. By extension the belowground structural complexity might influence diversity of subterranean organisms, but little work on this topic has been accomplished.

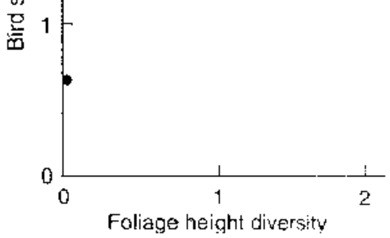


FIGURE 4 Relationship between foliage height and bird species diversity in areas of eastern North American deciduous forest. (Modified from MacArthur and MacArthur, 1961.)

The horizontal pattern of forest vegetation, particularly the arrangement of relatively uniform forest stand units across the landscape, influences biodiversity of the region and is the subject of a newly emerging branch of ecology—landscape ecology. The edges or ecotones (zones of rapid change in plant species composition) between forest stand units or forest associations often provide qualitatively different habitat than the interiors of those units; hence, the size, shape, and spatial arrangement of forest stands in the landscape influence the populations of associated species. Edges and sharp ecotones between stands may arise because of environmental discontinuities (e.g., topographic depressions where soil water accumulates) or because of the legacy of past disturbances. Thus, the local biodiversity of forests reflects both the compositional and structural diversity of the plants as well as the arrangement of units of relatively uniform composition and structure that we define as forest stands.

The structure of any particular forest stand traditionally has been defined as the distribution among age or size classes of the trees in a forest stand. Even-aged stands arise as the result of large-scale disturbances in which all or most of the large trees in an area are killed by a natural disturbance agent (e.g., crown fire, hurricane, fungal pathogen) or by human activity. The forest that arises on such a site consists of trees of roughly the same age, from new seedlings that colonize the site or from advanced regeneration (preexisting seedlings and saplings that escaped the disturbance). The process of stand development following large-scale disturbance results in gradual changes in the structure of the forest, as the growth, mortality, and recruitment of new individuals proceeds. These changes in forest structure include not only age structure, but also the spatial arrangement of the stems, branches, foliage, and roots of the plants that define the habitat of other biota in the forest. Moreover, changes in species composition of the forest usually accompany stand development, as species capable of growing in the shaded understory replace the pioneer species that colonize the disturbed site.

A typical sequence of forest stand development following large-scale disturbance has been characterized by Oliver and Larson (1990) as consisting of four stages: (stage 1) stand initiation, (stage 2) stem exclusion, (stage 3) understory reinitiation, and (stage 4) old growth. In the stand initiation stage, trees colonize the disturbed area. The time interval of this stage varies markedly depending on the severity of the disturbance, environmental factors at the site (e.g., climate and soils), and often herbivory. This stage concludes when the forest canopy becomes closed or when some soil resource (often water) becomes limiting to further increases in the leaf area of the forest. During the stem exclusion stage resources like light and soil water are so limited that suppressed understory trees die and regeneration is severely restricted. In this stage there is usually a continual reduction in the density (number of stems/area) of the original cohort of trees. The structure of the canopy is exceptionally simple at this stage as the individual trees grow in height to co-opt the light resource from neighbors.

In the understory reinitiation stage the overstory begins to break up as canopy trees die and the differential height growth of various species or individuals results in more complex arrangements of the foliage. Increased light reaching the understory favors the establishment and growth of new cohorts of species most capable of surviving in the highly competitive understory environment. The old growth or late successional forest stage is attained as overstory trees age and the canopy develops even greater complexity of structure. Gaps form in the canopy as a result of injury or death of the large, mature individuals, and the previously suppressed individuals are released from severe competition and grow in height. Decaying, coarse, woody debris accumulates on the ground and dead tree snags also provide new habitats for animals. Obviously, this idealized model of forest stand dynamics exhibits myriad local variations depending on the nature of the disturbance, environmental factors, and the tree species that dominate the area.

The changes in forest structure (e.g., foliage-height distributions) that accompany stand development following large-scale disturbance result in consequent gradual shifts in the quality of the habitat for different animals and plants. In the example cited earlier, MacArthur and MacArthur (1961) observed that maintenance of high numbers of bird species in eastern deciduous forests of the United States depended on the adequate provision of three layers of foliage, corresponding roughly to ground vegetation (0–2 ft), shrubs and small trees (2–25 ft), and overstory trees (>25 ft). If nearly all the foliage of the forest is in just one layer, as in the stand initiation and stem exclusion stages of stand development, bird diversity is much lower than in the old-growth stage, when canopy stratification becomes prominent.

The nature and degree of vertical stratification differ among forests. For example, temperate deciduous forests in the mature stage often exhibit the three strata just identified: an overstory stratum occupied by the canopy trees, an intermediate stratum represented by the crowns of saplings and understory species like dogwoods and hornbeams, and a ground stratum of low shrubs and forbs. In contrast, in the lowland tropical rain forest, very tall "emergent" trees with broad shallow crowns overtop the main canopy, which may be subdivided into two or more additional strata above the understory layers. And in the boreal forest the low stature and conical crowns of the conifers often preclude the formation of strong vertical stratification. These structural differences contribute to the contrasts in the number of distinct habitats provided for associated biota.

As noted earlier, many forest regions are only rarely affected by large-scale disturbances because they are both too moist to carry fires and not subjected to catastrophic windstorms. In these regions the steady-state forest is characterized as a shifting-mosaic landscape of small patches of different sizes and shapes, each patch reflecting the legacy of disturbance caused by the death of individuals or small groups of overstory trees. Each of the patches may follow a sequence of development analogous to that outlined for large-scale disturbances, but the overall structure of the forest is dependent on the arrangement of the tapestry of patches that comprise the larger forest stand. The edges between these patches and the vertical distribution of structural elements represent important dimensions of the habitat variability that permits species coexistence in forests.

Recognizing the importance of forest structure and pattern for biodiversity, ecologists are developing new, more sophisticated approaches for quantifying these parameters. Because the structural features that are important in determining animal and plant habitats differ among taxonomic groups and forest types, it is unlikely that any single approach will provide a universal standard by which forest structure and biodiversity can be related. Current efforts are utilizing new tools in the areas of spatial statistics, computer modeling, and remote sensing to provide suitable protocols for evaluating the connections between forest structure and pattern, management activities, and biodiversity of various groups of biota. These efforts will provide a better basis for understanding how forest ecology and biodiversity are related.

C. Food Webs and Community Organization in Forests

A food web is a set of species that live together and a specification of which species "eat" which other species. Plants provide the base of the food web by converting solar energy into biomass; herbivores and detritivores utilize living and dead biomass, respectively, to build their own tissues; and these organisms are in turn consumed by predatory species. Although a connection between the structure of food webs and biodiversity seems axiomatic, the exact nature of this connection is extremely complex. The biodiversity of a particular community is expressed in three different elements of food web structure: (a) the food chain length—the number of trophic links in the food web (producer-1°consumer-2°consumer-3°consumer etc.); (b) the number of distinct trophospecies—the set of all species that share some particular set of predators and prey—within a trophic level; and (c) the diversity of species constituting each trophospecies.

In forests, as distinct from most other biomes, the first link in the food web is dominated energetically by the detritivores rather than the herbivores. That is, most of the biomass in forests is consumed after the plant tissues die and are added to the soil as detritus. Thus, the detrital food web dominates the energetics of forest ecosystems and much of the complexity of the detrital food web in forests remains to be explored. These energetic considerations are not translated in a simple way to biodiversity in the respective food webs. The variety of different food sources available to herbivores in forests (leaves, stems, fruits, seeds, flowers, and roots of different plant species) provide numerous niches for their diversification, whereas most of the biomass energy available to detritivores is in the form of woody tissues which are structurally and biochemically so similar from species to species of trees that the diversification of wood decay organisms is somewhat limited. For example, one taxonomic group of insects, the higher termites (family Termitidae), overwhelmingly dominates in the comminution and decomposition of plant biomass in many tropical and warm temperate forest biomes. Nevertheless, in comparison with herbivore-dominated biomes (e.g., grasslands, aquatic ecosystems), the role of detritivores in forest biodiversity is probably relatively high.

The interactions of species in the forest community are not entirely competitive and predator-prey in nature. Some of the most fascinating interactions provide selective benefits to both of the interacting individuals or species populations—mutualistic and symbiotic relationships. Mutualism refers generally to a relationship in which two interacting species enhance their survival, growth, or reproduction, while symbiosis refers more specifically to two such organisms living together in close association. The most important symbiotic mutualism in forests is the mycorrhiza, an association between the mycelia of fungi and the roots of trees. The tree roots provide a supply of food to the fungus, which in turn increases the capability of the plant to acquire soil nutrients and water. All forest trees are mycorrhizal, and each tree species may harbor dozens of different fungal species in this mutually beneficial relationship. Two distinct types of mycorrhizae are common in forest trees—the ectomycorrhizae and endomycorrhizae. These types differ taxonomically, anatomically, and physiologically. The possible role of these mycorrhizal associations in regulating the diversity of forests was pointed out by Connell and Lowman (1989). They observed that pockets of low-diversity forest are found within the matrix of high-diversity tropical rain forest in all tropical regions. These low-diversity forests are composed of trees with ectomycorrhizal associations, whereas most of trees in the high-diversity forest are endomycorrhizal. Functional differences between the mycorrhizal types in soil nutrient acquisition or transfer could maintain or reinforce competitive interactions between individuals in these distinct forest types.

Most mutualisms between plants and animals have developed around the successful completion of the reproductive cycle of the plant—pollination and seed dispersal. Although many trees simply disperse their pollen to the wind, this method of pollination is unreliable when individuals are widely scattered in the forest, as in the species-rich tropical rain forest. Insects, nectivorous birds, and bats visit the plants to exploit them as a source of food and in the process carry pollen from one flower to another. Similarly, plants with seeds too heavy to be dispersed by the wind rely on animals to carry the seeds away from the mother plant. Although in most cases these mutualistic interactions are nonobligate and facultative (at least on one side), many remarkable examples of highly intricate, obligatory interactions have evolved, especially in the tropics. These interactions promote specialization and increased biodiversity.

III. HUMAN ACTIVITY AND FOREST BIODIVERSITY

Few forests have escaped the effects of human activity. Ancient civilizations decimated forests locally as a source of fuel and fiber and as sites for intensive agricultural production. In the modern era the pervasive influence of industrial civilization on forests has expanded to the regional and global scale through the additional effects of species introductions, air pollution, and likely climatic change. Insights from forest ecology provide a basis to evaluate the implications of these human influences on biodiversity.

A. Introduction of Alien Species

Alien insects and pathogens typically wreak havoc on host trees because these hosts have not developed adequate defenses through the process of evolution. The result is widespread decline of the host trees throughout their range even to near the point of extinction. The consequences of great reductions in the abundance of such declining species for the wider biotic community are not well understood and undoubtedly vary depending on the characteristics of the declining species (discussed later). Similarly, introductions of other species in different taxonomic groups or at other positions in the food web will have consequences for forest biodiversity that depend on their particular role in the community. Invasive trees may displace congeneric species from the forest community; and introduced herbivores that lack natural controls on their populations may decimate populations of their favored food plant species. Because the pace of species introductions has increased very rapidly in recent years, the ultimate consequence for forest ecosystems and biodiversity will be played out over the coming century.

B. Forest Harvest

The consequences of tree harvesting for forest biodiversity depend on particular features of the forest and the methods of harvest. Because all forests are regularly subjected to disturbance, if forest harvest practices mimic the natural disturbance regime, then consequences for biodiversity should be minimal. However, the exigencies of the financial bottom line result in harvest practices that do not mimic natural disturbances, and the consequences of actual forest harvest practices for biodiversity may be substantial. Most serious are (a) logging practices that result in the failure of the cut-over site to regenerate (e.g., because of severe damage to soils); (b) the coincident harvest of extensive areas, so that most of the landscape is in a single stage of forest stand development; and (c) recurring harvest on short rotation intervals. Also, forest harvest differs fundamentally from natural disturbance in that wood products are removed from the site; any species that depends on decaying wood for its habitat will be harmed by harvest practices that do not recognize this dependency. Finally, some species are believed to be old-growth obligates (i.e., they depend on old forests to complete their life cycles). These species are threatened when the great majority of natural, old forests in a region enters the harvest-regrowth system of industrial forestry, leaving little old-growth habitat.

C. Forest Conversion and Fragmentation

Permanent or semipermanent conversion of forested areas to other land uses has more severe consequences for biodiversity than forest harvest. Many forests occur where climate and soils are suitable for permanent agriculture and where the expansion of urban communities gobbles up native vegetation. Maintenance of biodiversity in such regions depends on having protected forest areas large enough to harbor the native flora and fauna. However, general rules to guide forest preserve planning for biodiversity protection are complicated by variations in the habitat requirements of different species. Forests in most agricultural regions occur as small fragments dispersed across the landscape, and how effectively these fragments can maintain viable populations of forest species is a topic of great concern (Schelhas and Greenberg, 1996).

D. Pollution

Local declines of forests has been associated conclusively with point-source releases of air pollutants, especially sulfur dioxide, fluoride, and toxic metals from smelters. Broad-scale, regional effects of air pollution on forests have been more difficult to demonstrate. Regional pollution—by ozone smog in the southwestern United States and by acidic deposition in the eastern United States and Europe—probably has contributed to documented forest declines. Although improvements in emission controls and regulations in these regions are likely to reduce the chances of further damage, rapid industrialization without adequate emission controls in other regions of the world threatens forest health and biodiversity.

E. Rapid Climatic Change

The Pleistocene epoch was marked by dramatic climatic shifts that profoundly affected forests and biodiversity. The rapid rise in greenhouse gas concentrations is likely to bring about similar climatic shifts in coming decades or centuries. Of course, the consequences of rapid climate change for forests and biodiversity will depend on a combination of species' natural responses (e.g., dispersal, colonization, natural selection) and human mitigation efforts. In many forest regions certain species, like the dominant trees and particular wildlife populations, are likely to be controlled by management efforts because of their relatively high value to humans. For relatively low-valued species and forest regions, the maintenance of biodiversity may depend on natural mechanisms or heroic human efforts that recognize nonmarket values of species. The current level of understanding of the physiological and population ecology of many forest-dwelling taxa is insufficient to predict the effects of rapid climatic change, but species with limited capacity for dispersal and colonization (e.g., soil invertebrates, perennial herbs) may be most sensitive. The implication of the loss of these species from forest communities will vary depending on the role they play in forest ecosystem function.

IV. BIODIVERSITY AND FOREST ECOSYSTEM FUNCTION

Forests regulate energy flow and cycling of materials in the landscape, collectively known as ecosystem functions. The effects of biodiversity, expressed in terms of species richness, on forest ecosystem functions are not yet clear and apparently not very straightforward. There is great interest in these possible effects as ecologists probe the implications of loss of species diversity for the integrity of ecosystems. Will species extinctions result in destabilization of ecosystem functions and possible feedbacks in the form of undesirable shifts in dominant vegetation types?

Whereas the role of species richness per se in regulating forest ecosystem functions remains unclear, it is well known that loss of particular species from the biota of a community can have important ramifications for energy flow, material cycling, and maintenance of stable biotic composition. That is, all species in the ecosystem are not equal in terms of their quantitative influence on ecosystem function. In particular, for some species there appears to be little or no redundancy with respect to their role in the ecosystem; if such a species performs some crucial activity, its loss from a forest can create havoc for the normal functioning of the ecosystem. These species are known as keystones.

In forests keystone species are represented among many different taxonomic groups or food web positions. For example, at the primary producer level, a nitrogen-fixing tree like *Alnus rubra* in N-poor conifer forests is a keystone species; elephants appear to be keystone herbivores in semiarid Africa; beavers are keystone "ecosystem engineers" in northern forests; and jaguars that prey on seed predators in neotropical forests may be keystone carnivores. The loss of these species has consequences for ecosystem structure, function, and composition that are out of proportion from their individual abundances. Important efforts to conserve forest ecosystem functions in the face of biodiversity loss are focused on the identification of keystone species and ways of maintaining stable populations of keystone species.

Although the broader effects of overall reductions in biodiversity on forest ecosystem function remain more obscure, a specific example will illustrate that this is also a cause for concern. During the past several decades, excessive inputs of nitrogen (from air pollution) to forests in northern Europe have resulted in striking reductions in the abundance of mushroom species. These mushrooms are the fruiting bodies of ectomycorrhizal fungi discussed earlier. Lilleskov *et al.* (2000) have shown that a single host tree, white spruce (*Picea glauca*), maintains associations with about 90 different mycorrhizal fungi in natural, N-poor forests in Alaska, whereas in adjacent N-polluted forests only about five fungal associates are found. The high diversity of this mycoflora in the natural forest certainly represents some degree of functional redundancy. However, if dozens of species are lost from the mycorrhizal fungal flora in temporarily N-polluted regions, some long-term effects on forest ecosystem function are likely. The fungi in the N-rich forests appear to utilize only the mineralized nitrogen sources (NH_4^+ and NO_3^-), which are abundant there; with a return to normal, low–mineral N availability, in the absence of mycorrhiza that access organic N forms, the productivity and nutrient cycling in the forests could be altered profoundly. Analogous situations probably apply in other aspects of both the detrital and grazing food webs of forest ecosystems around the world.

See Also the Following Articles

BOREAL FOREST ECOSYSTEMS • DEFORESTATION AND LAND CLEARING • DISTURBANCE, MECHANISMS OF • FIRES, ECOLOGICAL EFFECTS OF • FOOD WEBS • KEYSTONE SPECIES • LOGGED FORESTS • RAINFOREST LOSS AND CHANGE • REFORESTATION • SUCCESSION, PHENOMENON OF

Bibliography

Barnes, B. V., Zak, D. R., Denton, S. R., and Spurr, S. H. (1998). *Forest Ecology*, 4th ed. John Wiley & Sons, New York.

Connell, J. H., and Lowman, M. D. (1989). Low diversity tropical rainforests: Some possible mechanisms for their existence. *American Naturalist* 134, 88–119.

Gentry, A. H. (1995). Diversity and floristic composition of neotropical dry forests. In *Seasonally Dry Tropical Forests* (S. H. Bullock, H. A. Mooney, and E. Medina, eds.), pp. 146–194. Cambridge University Press, Cambridge, England.

Latham, R. E., and Ricklefs, R. E. (1993). Continental comparisons of temperate-zone tree species diversity. In *Species Diversity in Ecological Communities* (R. E. Ricklefs and D. Schluter, ed.), pp. 294–314. University of Chicago Press, Chicago.

Lilleskov, E., Fahey, T. J., and Lovett, G. M. (2000). Ectomycorrhizal fungal community change over an atmospheric nitrogen deposition gradient in Alaska. Ecological Applications, in press.

MacArthur, R. H., and MacArthur, J. W. (1961). On bird species diversity. *Ecology* 42, 594–598.

Oliver, C. D., and Larson, B. C. (1990). *Forest Stand Dynamics*. McGraw Hill, New York.

Schelhas, J., and Greenberg, R. (1996). *Forest Patches in Tropical Landscapes*. Island Press: Washington, DC.

Whitmore, T. C. (1998). *An Introduction to Tropical Rain Forests*. 2nd ed. Oxford University Press, Oxford, England.

FOSSIL RECORD

Sean R. Connolly
University of Arizona

GLOSSARY

chronostratigraphic correlation Matching strata in different locations according to the points in time to which their boundaries correspond.

extant Having some living representatives (i.e., not extinct) at a particular time.

molecular phylogenetics The study of the history of evolutionary relationships according to similarities and differences among (usually extant) organisms in molecular characters (such as DNA).

parametric statistics Statistical methods that assume data have certain characteristics, such as a particular kind of distribution (Poisson, normal, binomial, etc.).

stratigraphic interval A stratum with identifiable upper and lower boundaries.

stratigraphic range The strata through which a taxon is known to have been extant.

stratum (plural: strata) An interval of sedimentary rock that is distinguishable from previous and subsequent intervals.

taxon (plural: taxa) A group of organisms that share a name.

OBTAINING RELIABLE ESTIMATES OF biological diversity in the fossil record is critical to addressing key questions about changes in biological diversity over long timescales and the causes and consequences of those patterns. On what scales is the history of life characterized by periods of stasis punctuated by relatively brief periods of rapid evolutionary change? On what scales is it a more gradual, continuous process? What are the causes of these changes? How often are major changes in biodiversity associated with major changes in the physical environment, such as extraterrestrial impacts? Unfortunately, the fossil record provides an incomplete picture of the history of biological diversity. This chapter reviews some important features of the fossil record that could cause the apparent history of biodiversity to differ quite markedly from its true history. It then introduces several methods that have been applied in an attempt to account for these features, it discusses advantages and disadvantages associated with each of them, and it reviews how their application in specific cases has altered earlier views about changes in diversity through time. Throughout this chapter, the terms "biodiversity" and "biological diversity" will refer to taxonomic richness: the number of *taxa* present at a particular time or place. Although alternative metrics of diversity have been proposed, most paleobiologists have focused on taxonomic richness, at least when addressing regional or global patterns in biodiversity.

I. SAMPLING EFFECTS IN THE FOSSIL RECORD

A. The Fossil Record Is Incomplete

One obvious feature of the fossil record is that it is incomplete: only a small fraction of individuals are fossilized; of those, very few are collected and identified. As a result, the biological diversity recorded in the fossil record is less than total diversity over the region and time interval from which it is a sample. An additional consequence of this fact is that it makes stratigraphic ranges of taxa correspond to shorter periods of time than their true life spans. This is because the first appearance of a taxon in the fossil record occurs sometime after it actually originated, unless its very first representative was fossilized and subsequently sampled (a highly unlikely prospect). Similarly, the last appearance of a taxon occurs sometime before it actually went extinct. Biased estimates of stratigraphic ranges are important for the study of trends in biodiversity because they are often used to estimate biodiversity. Thus, rather than estimating diversity as the number of taxa that actually occur in samples from a particular stratigraphic interval, diversity is estimated as the number of taxa whose stratigraphic ranges encompass that interval. This will always be equal to or greater than the number of sampled taxa.

B. The Quality of Preservation Varies

If incomplete sampling were the only problem with the fossil record, it would still be possible to accept relative trends through time at face value—that is, an increase in diversity in the fossil record would indicate a real increase in biological diversity, even if the true number of extant taxa could not be determined. Unfortunately, however, the degree to which the fossil record is incomplete varies in both space and time. One source of this variation is characteristics of organisms themselves. For instance, organisms with hard parts are much more likely to be preserved in the fossil record than those without them. Clams, for instance, have a better fossil record than nematodes. One consequence of this is obvious: the difference between apparent and true diversity tends to be greater for the latter group than for the former. However, estimates of diversity trends through time can be affected as well, if the proportion of taxa with body parts that are readily fossilized does not remain constant. Indeed, some have argued that the "Cambrian explosion" represents, not an explosion of multicellular life, but a rapid and extensive proliferation of hard parts. Recently, this hypothesis has been fueled by molecular phylogenetic studies predicting that the major animal phyla diverged long before the early Cambrian. This hypothesis remains highly controversial; nevertheless, the very fact that it has received considerable attention illustrates just how profoundly sampling effects are believed to influence the fossil record of major events in the history of life.

Differences in abundances among taxa can also affect their preservation in the fossil record. It is individuals (or parts of individuals) that are fossilized; thus, more abundant taxa are, on average, likely to have more complete fossil records than will rare taxa. Further, fossil diversity is likely to be higher for intervals during which abundances were higher, on average, than during other intervals. This, too, has important consequences for inferences about diversity trends. For instance, several major episodes of diversification, as recorded in the fossil record, coincide with geophysical changes that probably increased rates of nutrient supply to the biosphere. Some workers (e.g., Vermeij, 1995) have argued that these geophysical changes were important causes of the coincident biological diversification. However, if increases in rates of nutrient supply also allowed taxa to sustain higher abundances, then increased probabilities of sampling during those sequences could be contributing to the increased diversity of the fossil record. Thus, inferences about the causes of trends in biodiversity, in addition to inferences about the trends themselves, can be influenced by sampling effects.

Differences in abundances can also affect estimates of stratigraphic ranges. One of these effects is illustrated by the previous example: first appearances of new taxa in the fossil record will be higher between two stratigraphic intervals when large increases in abundance occur between them as well, even if there were, in fact, an equal number of originations during the time periods that correspond to those intervals. Assessing the severity of this effect is complicated by the fact that there are sound biological reasons for increases in abundance to facilitate originations of new taxa; to the evolutionary biologist, a correlation between abundance and rate of speciation may be precisely what is expected!

Yet another characteristic of organisms that can affect inferences about biodiversity trends is habitat. In particular, individuals in some habitats are more likely to become fossils than individuals in others. For instance, marine soft-bottom habitats are likely to provide more complete records of their inhabitants than will rocky shores. In the former case, individuals will much more readily be covered in sediment shortly after (or even before) they die. In the latter, wave action may render the remains of individuals unidentifiable before currents carry them to a location where they might be buried by sediment and preserved. Indeed, the fossil record of rocky shore communities is among the poorest in the marine realm. Again, the difficulties associated with these effects has led to disagreements about key events in the history of life. For instance, fossils record an explosive diversification of mammals in the early Tertiary, following the mass extinction that ended the Cretaceous period (best known for catastrophic extinctions among dinosaurs). Most workers believe that this reflects the true pattern, at least qualitatively. However, others, using molecular phylogenetics, have argued that much of this diversification occurred in the Cretaceous, prior to this mass extinction. One explanation for this discrepancy has been that these early Cretaceous mammals occupied habitats (such as forest interiors) for which their probability of being preserved was much lower than for their Tertiary descendants, who expanded into new habitats to occupy niches vacated by the extinction of dinosaurs. Whether this discrepancy between molecular and fossil data is primarily due to a poor fossil record for Cretaceous mammals or to shortcomings in the molecular methods remains highly controversial. Attempts to assess the severity of these shortcomings, such as those of Foote and coworkers (1999), should ultimately lead to a resolution.

As this discussion of the effects of ecology suggests, the probability of an individual being preserved in the fossil record depends, in part, on the sediments themselves. This, too, can have profound consequences on inferences about diversity patterns. For instance, in a classic paper, Raup (1976) noted a systematic increase in the volume of sedimentary rock through time. The implication was that the probability of individuals being preserved in the fossil record becomes progressively greater through time. Indeed, he presented a graph of sedimentary rock volume through time that looked strikingly similar to a graph showing biological diversity in the fossil record through time. After removing the effect of rock volume on diversity, he found no evidence for a long-term increase in species diversity through time. Although a key paper by Sepkoski *et al.* (1981), along with subsequent work, has convinced most workers that the apparent long-term diversity increase is real; the extent to which systematic increases in rock volume exaggerate the trend remains unknown.

This long-term increase in sedimentary rock volume has another potential effect on large-scale diversity patterns. If the probability of preservation and sampling increases through time, then the difference between the time of origination of a taxon and the time of its first appearance in the fossil record should shrink. Similarly, the difference between the time of extinction and the time of last appearance should shrink. This effect would lead to a progressive decrease in apparent extinction rates through time. Indeed, a long-term trend of decreasing extinction rates has been noted in work in the 1980s by Sepkoski and Raup and by Van Valen. Knowing the extent to which this trend reflects a real decrease in extinction rates (rather than a sampling effect as just described), is important, because Sepkoski (1984) has noted that such a trend, if real, could explain long-term changes in the relative prevalence of different taxa in the fossil record. As a result, paleobiologists (e.g., Pease, 1988, 1992) have proposed methods for assessing the effects of sampling biases on these rates. There appears to be an emerging consensus that the declines in evolutionary rates are real, but that they may be exaggerated by coincident trends in the quality of preservation.

The extent and quality of the fossil record varies over smaller timescales as well. Geophysical transitions can dramatically affect the probability that individuals will be preserved as fossils. When these transitions are global in scope, or occur in regions that receive disproportionately large amounts of attention from paleontologists, then apparent changes in diversity can result. Just such a possibility has complicated analysis of what was probably the most extensive mass extinction in the history of life, an event that marks the end of the Permian and the beginning of the Triassic. This event coincided with a major loss of marine benthic habitat (due to decreasing sea level). While it is certainly reasonable to suspect that biodiversity would decrease as available habitat decrease, it would also reduce the probability of sampling taxa that remained, as noted by Signor and Lipps in an influential paper published in 1982. This and other sampling effects associated with characterizing this extinction event and its aftermath are extensively reviewed by Erwin (1993).

C. The Extent of Sampling Varies

Decisions paleontologists make can also introduce, or exaggerate, differences in the quality of the fossil record. For instance, a disproportionately large amount of effort has been devoted to sampling in the Mesozoic and Cenozoic eras, because paleontological sampling tends to be associated with petroleum exploration, and rocks that date from these areas have been subject to more extensive exploration than older rocks. Similarly, North America has been more extensively sampled than many other regions of the world. This reflects, in part, the fact that many paleontologists live in North America. The logistical problems associated with fieldwork are generally less severe, and the cost of travel lower, when fieldwork is conducted near a home institution. In addition, fieldwork tends to be easier in regions with a well-developed infrastructure. As a result, more of the preserved biodiversity is actually sampled and recorded, and thus the apparent diversity of well-sampled regions is higher than that of poorly sampled regions.

Similarly, the questions that one can ask of paleontological data, and the degree of confidence that can be assigned to answers, depends in large measure on the quality and quantity of data that can be collected. As a result, disproportionately large numbers of paleontologists focus their efforts on taxa, stratigraphic intervals, and geographical regions where fossils are abundant and well-preserved. This tends to exaggerate the effects of differences in preservation: where preservation is good (something which, alone, would tend to make biodiversity appear greater), a larger proportion of that record has been sampled. One effect of this tendency is that more workers focus on the most recent era of time, the Cenozoic (0–65 million years ago), than on the previous era, the Mesozoic (65–250 million years ago), despite the fact that the latter era was longer in duration. Likewise, more paleontologists study the Mesozoic than the earlier Paleozoic (250–600 million years ago).

D. Time-Averaging

Yet another factor affecting estimates of diversity patterns is time averaging. When fossils are collected, their location in time is generally reported according to the particular stratum in which they were found. This corresponds to an interval, rather than a specific point, in time. Some of these strata represent longer periods of time than others. Thus, the number of taxa found in a particular stratum may be greater than the number of those taxa that were actually extant at a particular point in time during that interval (assuming that extinctions or originations occurred during the interval). The longer the interval lasts, and the higher the rates of origination and extinction, the more apparent diversity is inflated.

The issue is further complicated by uncertainties associated with the points in time that correspond to the boundaries of stratigraphic intervals. Indeed, the ages of these boundaries are continually revised. Thus, it may not always be clear which intervals were longer than others, and it is never clear just exactly how long those intervals were. Unfortunately, this problem becomes more severe as intervals are more finely divided. This makes minimizing the time-averaging problem more complicated. That is, the time-averaging problem should become smaller as the length of time spanned by an interval decreases. Since stratigraphic intervals are classified hierarchically (with some types of intervals nested within larger ones), one could, in principle, minimize the time-averaging problem by using a low level in this hierarchy. Unfortunately, however, the shorter the interval, the greater the proportional uncertainty associated with the estimated ages of the intervals' boundaries. For instance, overestimating the age of the lower boundary of a 10 million year interval by 1 million years inflates the estimated duration of that interval by 10%. However, an overestimate of the same magnitude on the duration of a 3 million year subunit of that interval represents a 33% inflation.

In addition, it can be difficult to determine how stratigraphic intervals in one location correspond to intervals in other locations. If one is interested in estimating diversity or macroevolutionary rates for a region that includes more than one sampling location, then this can introduce further uncertainties. One source of this uncertainty is a simple extension of the problem of uncertainty in dating the boundaries of particular strata: if there are errors in the estimates of the ages of these boundaries at two different locations, then two intervals may be assumed to be substantially coincident when, in fact, they are not. This problem, like that discussed in the previous paragraph, becomes more severe as the duration of stratigraphic intervals decreases.

A second problem associated with this process of chronostratigraphic correlation results from the fact that direct estimates of ages are not available for all stratigraphic boundaries in all locations. Often, correlations are based on the presence or absence of particular indicator taxa. That is, the stratum in which a particular taxon first appears at, say, location B is assumed to correspond to the stratum in which it appears at location A. If the ages of the boundaries of the stratum at location A have been estimated, but those of location B have not, then the lower boundary of the stratum at location B is assumed to fall within that range of ages. However, since new taxa originate in particular regions, then expand their ranges gradually into new regions, part of the uncertainty associated with this effect is based on this rate at which indicator taxa increase their geographical ranges. In addition, however, as noted earlier, the difference between the time at which a taxon was first present at a location and the time of its first appearance in the fossil record for that location varies depending on factors specific to particular locations, such as quality of preservation and local abundance. This further contributes to uncertainties in stratigraphic correlation. In practice, paleontologists attempt to minimize this problem by using multiple indicator taxa. They also emphasize those taxa that are likely to have high probabilities of preservation in the fossil record and whose spread is likely to have been rapid, such as planktonic foraminifera.

II. ESTIMATING DIVERSITY TRENDS

A. Rarefaction

One way to control for these sampling effects in making inferences about diversity trajectories is to compare taxonomic richness among locations or strata with samples that are equivalent in extent. This technique, called *rarefaction*, was developed by Sanders (1968)—and amended by Simberloff (1972)—to compare the diversities of different habitats in present-day ecosystems. The approach is as follows: a sample is collected from each of a set of habitats (for instance, a certain volume of sediment is obtained) using an identical sampling scheme. All of the individuals in each sample are identified, and their taxonomic identity is recorded. The next steps can be visualized by imagining placing all of the records for a particular sample in a bowl, stirring them, and then randomly selecting records from that bowl until none are left. Each time a record is picked, the taxonomic identity of that sample is noted. From this sequence of random draws, one constructs a graph with the number of records on the horizontal axis and the number of distinct taxa on the vertical axis. What is plotted on that graph are the results of that sequence of record selection. This curve must intersect the points (0, 0) and (1, 1). That is, before the experiment begins, 0 records have been picked from the bowl, so 0 distinct taxa have been selected. When one record is chosen, exactly one distinct taxon has been found. When two records have been chosen, either one or two distinct taxa have been found. [If the second record has the same taxonomic identity as the first, the next point is (2, 1); if it has a different identity, the next point is (2, 2)]. Once all of the points have been plotted, a curve is fit to them: this is the rarefaction curve. This process is repeated for each sample collected, so that each sample has a rarefaction curve associated with it. Biodiversity in the different samples is then compared by examining the number of distinct taxa encountered for a given number of records selected. That number must be equal to or less than the number of records in the sample with the fewest records. This comparison is illustrated in Figure 1.

Figure 1 illustrates something else: rarefaction curves sometimes cross. This means that the rank order of diversities of a set of samples can change, depending on sample size. One might therefore ask what rarefied diversity means, and, further, what it means to say that one site has a greater rarefied diversity than another. Hurlburt (1971) ventured an answer to that question, proposing that rarefied diversity measured the number of distinct taxa encountered, on average, by an individual organism in a particular habitat over the course of a particular number of encounters with other individuals (provided that individuals are not encountered multiple times). The fact that one habitat has a greater rarefied diversity than another for a given number of occurrences of individuals does not mean that that habitat actually has a greater taxonomic richness (i.e., contains more taxa) than another.

Although rarefaction can be applied to the fossil record in the manner just described, it is often applied somewhat differently. When comparing diversities over very large scales (global diversity during two strati-

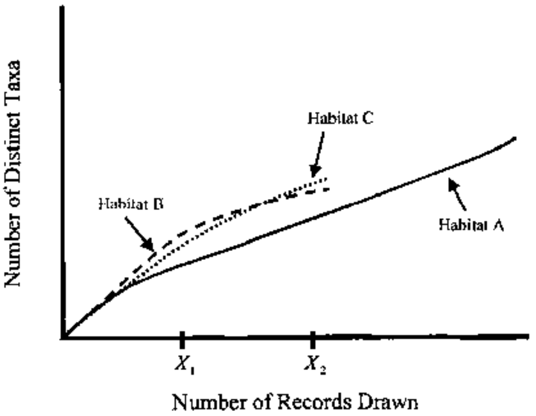


FIGURE 1 Hypothetical rarefaction curves for three samples representing three different habitats, designated A (solid line), B (dashed line), and C (dotted line). Note that habitat A, whose sample contains the most species in its entirety (note that its rarefaction curve ends at a higher diversity than the others), is not the most diverse habitat when its diversity is compared with the others for a particular sample size (X_1 on the horizontal axis). Also note that the rarefaction curves for samples B and C cross. Thus, habitat B is more diverse when there are X_1 records in the sample, but habitat C is more diverse when there are X_2 records.

graphic intervals, for instance), a sample enumerating all of the individuals in those intervals will not be available. Rather, data will consist of a number of distinct samples for each of the regions or intervals being compared. Further, while a list of the taxa found in each sample may be available, information on the abundances of those taxa may not. Such data have been rarefied in two different ways. In one approach (hereafter called *rarefaction by occurrence*), a rarefaction curve is constructed for a habitat by randomly selecting occurrences of taxa from its associated samples. The rarefaction curve is then a plot of the number of distinct taxa sampled against the number of occurrences of taxa drawn from the available samples. This is illustrated in Figure 2. Diversities are compared by constructing these rarefaction curves for different regions, taxa, or stratigraphic intervals. An alternative approach (hereafter called *rarefaction by list*) involves randomly selecting the entire list of taxa present in a given sample as a unit, rather than selecting occurrences of taxa within those samples. In this case, the rarefaction curve for a particular habitat, region, or stratum consists of the number of distinct taxa detected plotted as a function of the number of samples drawn. Note that in this instance, the curve is not constrained to pass through the point (1, 1), since several taxa may appear in a single sample. However, it still must pass through the point (0, 0). This method is illustrated in Figure 3.

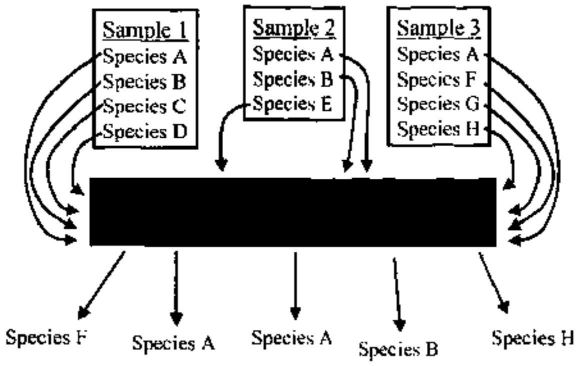


FIGURE 2 A graphical illustration of rarefaction by occurrence. Occurrences of a taxa within samples are treated independently, and these occurrences are drawn individually at random from the set of occurrences in all samples. This can be visualized as adding each occurrence individually (regardless of the sample from which it comes) into a pool (arrows into the black box), then randomly choosing these occurrences from that pool (arrows out of the black box). In this figure, the rarefied diversity after five records have been drawn is four distinct species.

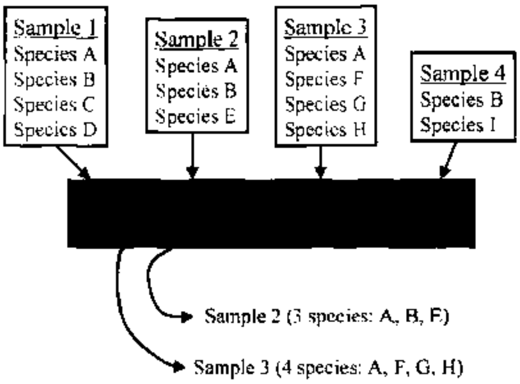


FIGURE 3 A graphical illustration of rarefaction by list. In this case, an entire paleontological sample is drawn at random from the population of samples that constitute the data set, and rarefied diversity is the total number of unique species contained in those lists for a given number of samples drawn. In this figure, the rarefied diversity after two samples have been chosen at random is six species (A, B, E, F, G, and H).

Rarefaction has been applied to paleontological data in several instances. Raup (1975) pioneered its use in paleobiology. He examined an apparent increase in diversity of echinoid families since the Paleozoic. Through rarefaction, he demonstrated that this increase was qualitatively unchanged after accounting for increased sampling in younger stratigraphic intervals. In the 1990s, Miller and coworkers applied rarefaction by occurrence to explore taxonomic and regional patterns in the diversification of benthic marine invertebrates during the Ordovician Radiation. Unlike Raup, they found that broad-scale patterns suggested by raw data were changed by rarefaction. For instance, global rarefied diversities increased only through the mid-Ordovician, when they stabilized. This contrasts with the trend apparent from the fossil record as a whole: increasing diversity through the late Ordovician. Similarly, large peaks in diversity at the scale of individual paleocontinents were often reduced or eliminated by rarefaction. By contrast, Alroy (1999) applied rarefaction by list to explore the extent to which the radiation of mammals preceded the end-Cretaceous mass extinction (see Section I.B, "The Quality of Preservation Varies"). His results agreed qualitatively with the trend apparent from the "uncorrected" fossil record: much of the taxonomic diversification of mammals appeared to occur after this mass extinction.

Paleontologists have typically been very cautious in their interpretation of rarefied diversities. There is a reason for this caution: the biological meaning of a rarefied diversity measurement is unclear in the paleobiological context. The interpretation offered by Hurlburt (described earlier) is appropriate only when (a) the sampling protocol described in the first paragraph of this section has been followed or (b) each occurrence of a taxon is a truly independent, random sample from the habitat about which one wishes to make inferences. Generally, neither is true for large-scale paleobiological diversity estimates, because sampling effort has not been distributed randomly. For instance, when comparing the rarefied diversities of two stratigraphic intervals, a disproportionately large number of samples may be available for one habitat or region during one interval, but not for the other interval (but see methods developed by Alroy [e.g., 1998] for minimizing these biases). In addition, for rarefaction by occurrence, occurrences of taxa are not independent, because some occurrences come from the same sampling locations and others from different sampling locations. A biological interpretation of rarefied diversity given these sampling effects remains elusive. Nevertheless, rarefaction can provide important information. In particular, discrepancies between rarefied and overall diversity patterns indicate where marked increases or decreases in apparent diversity are likely to be artifacts of sampling.

B. Capture-Recapture Estimates

Although rarefaction has probably been the most widely used tool to account for the effects of variation in sampling effort, other approaches have been tried as well. For instance, in the mid-1980s, Nichols and coworkers proposed that models used to estimate abundances from capture-recapture data in population biology could be adapted, by analogy, to estimate taxonomic diversity in the fossil record. When one conducts a capture-recapture study, individual organisms are captured during discrete sampling occasions, which can occur at different locations, different times, or both. Each individual captured is given a unique mark, so that its capture history can be constructed. That is, if one constructs a matrix, the columns j of which represent different sampling occasions and the rows i of which correspond to each individual captured at least once, then each element a_{ij} in the matrix will be either 1 or 0, indicating whether individual i was captured on occasion j. Similarly, fossil data are collected on discrete sampling occasions, and one can simply list the taxa sampled on each occasion. In this case, the matrix elements a_{ij} are 1 or 0 according to whether taxon i was found in sample j. In this context, taxonomic diversity (the total number of distinct taxa) is analogous to population size (the total number of distinct individuals). Unlike rarefaction, these approaches estimate taxonomic diversity (rather than sampled diversity given a particular sample size) when the models' assumptions are met. Unfortunately, however, these assumptions are rarely fully met by fossil data. It is not always clear which results are robust to violations of assumptions and which are not. Therefore, as with rarefaction, workers have been cautious when interpreting these estimates.

Two different types of capture-recapture models can be used to estimate diversity, depending on the sampling scheme. When sampling occasions occur at different times and the duration of a particular sampling occasion is short relative to the time between sampling occasions, open-population models can be used. These models are designed to account for "births" and "deaths" of taxa (i.e., originations and extinctions) that may occur between sampling occasions. In the paleobiological context, stratigraphic intervals have been treated as sampling occasions. Thus the taxa sampled in an interval are considered as having been "captured" at some point during that interval (usually its midpoint).

By contrast, when sampling occasions are sufficiently close together that originations and extinctions are few relative to the total number of taxa extant, then closed-population models can be used. These models have been applied when multiple samples have been taken within a stratigraphic interval. Although these models assume that there are no originations or extinctions between the specific points in time represented by different samples, they tend to be more robust to other types of sampling problems (discussed later). The particular models that have been applied to paleontological data include the Jolly-Seber open-population model and the closed-population models of Burnham and Overton (1979) and Chao (1987).

1. Jolly-Seber Model

The Jolly-Seber model makes four key assumptions:

1. The strata that represent different sampling occasions are short in duration relative to the time between those strata.
2. All taxa have the same probability of being sampled within a particular stratum.
3. All taxa have the same probability of going extinct between strata j and $j + 1$.
4. If a taxon goes extinct in the region being sampled, it does not subsequently reinvade that region from elsewhere.

Given these assumptions, diversity (D_i) can be estimated for any stratum i from the following data (notation follows that used by Nichols and Pollock, 1983): the number of taxa sampled during i (n_i), the number of n_i that were also sampled during at least one earlier interval (m_i), the number of n_i that were also sampled during at least one later interval (r_i), and the number of taxa sampled at least once before i, at least once after i, but *not* during i itself (z_i). The first step in estimating diversity involves estimating the number of taxa extant, but not sampled, during i (M_i). Note that r_i/n_i is the proportion of taxa sampled during i that were sampled again during a later interval. Similarly, $z_i/(M_i - m_i)$ is the proportion extant but *not* sampled during i that were sampled again subsequently. If the model's assumptions hold, these two fractions will be equal, and M_i can be estimated. The second step involves noting that m_i/n_i is the proportion of taxa sampled during i that were sampled during earlier intervals. Similarly, M_i/D_i is the proportion of taxa extant during i that were sampled during earlier intervals. Again, if the model's assumptions hold, these two fractions will be equal, and diversity can be estimated.

As discussed earlier, taxa differ in many characteristics that affect their probability of being sampled as fossils (e.g., presence of hard parts, habitat, abundance). This violates assumption (2) and causes estimated diversity to tend to be lower than true diversity. Perhaps not surprisingly, then, when Nichols and Pollock (1983) applied this model to late Eocene mammals from the Big Horn Basin, Wyoming, goodness of fit statistics indicated rejection of the model. When applied to molluscan diversity in the Middle Miocene of South Jutland, the model provided an adequate fit. However, diversity estimates were lower than those obtained with a method that does not assume equal sampling probabilities (discussed later).

2. Closed-Population Models

Several alternative models can be used when multiple samples are available for a particular time interval. Those that have been applied to the fossil record share one important feature in common: they are designed to allow for the possibility that some taxa are more likely to be sampled than others. All of these methods utilize the frequency distribution of occurrences of the taxa sampled. That is, the raw material for the diversity estimate is the number of taxa occurring in only one sample, f_1, the number occurring in two samples, f_2, and so on, as well as the total number of occurrences in all samples.

Burnham and Overton (1979) utilized a statistical approach known as the jackknife to estimate biodiversity. The mathematics of the derivations are too complex to review here, but with this approach they obtained a series of possible estimators. The simplest of these utilizes only the number of samples and the number of taxa occurring in only one sample:

$$\hat{D}_1 = D_{obs} + f_1 \frac{k-1}{k} \tag{1}$$

where $\hat{D}_1$ is the *first-order jackknife* estimate of diversity, D_{obs} is the number of distinct taxa appearing in the sample, f_1 is the number of taxa occurring in exactly one sample, and k is the number of samples. They developed additional estimates by incorporating the number of taxa occurring in more than one sample (for instance, their *second-order jackknife* uses the number of taxa appearing in exactly two samples, as well as the number appearing in just one). Nichols and Pollock (1983) applied these models to same molluskan data on which they used the Jolly-Seber method described earlier. They found that, even under a relatively intensive sampling regime, sampled diversity was as much as 30% lower than estimated diversity.

An alternative model, proposed by Chao (1987), uses the number of taxa occurring in either exactly one or exactly two samples:

$$S_2 = D_{obs} + \frac{f_1^2}{2f_2} \tag{2}$$

Wing and DiMichele (1992) used this estimate, usually termed "Chao-2," to compare regional vegetation diversity in the late Paleozoic and early Cenozoic for North American river and delta floodplains. Somewhat surprisingly, they found similar biodiversity levels during the two periods for these regions, despite the markedly higher apparent diversities for the latter interval at the global level.

Like the other methods discussed, these models make several assumptions that limit their applicability. The jackknife and Chao estimates assume that each taxon has an equal probability of occurring in each sample. That is, taxon A may have a different probability of being present in a sample than taxon B, but that taxon-specific probability is the same for every sample within each region. This assumption may be violated if samples differ in extent or quality. However, even if great care is taken to minimize this problem, the assumption may still be violated. For instance, if abundances of particular taxa differed among sampling locations, then the associated probabilities of sampling may differ accordingly.

C. Generalized Inverse Gaussian-Poisson Distribution

Models other than those based on capture-recapture theory utilize frequency distributions of occurrences to estimate diversity. One of these methods is known as the Generalized Inverse Gaussian-Poisson (GIGP) distribution. This method involves fitting a statistical distribution of specified form to the observed frequency distribution, then extrapolating from this statistical distribution to estimate the number of species that have been sampled zero times, f_0. Total biodiversity, then, is simply this number plus the number of species that were sampled at least once. Anderson and coworkers (1996) applied this method to plant and insect data for late Triassic braid-river plains in South Africa. They used these estimates to argue that plant and insect diversity in the sampled habitats was comparable to those of the present day, again in contrast to apparent global diversity patterns, which record increasing biodiversity levels through time.

The primary limitation of this approach is that it assumes that the underlying frequency distribution of occurrences follows a particular statistical distribution. That is, it is a parametric method. In this respect, it differs from the closed-population capture-recapture methods discussed earlier, which are nonparametric. As a general rule, parametric methods are more precise than nonparametric analogs when the assumptions about the underlying distribution are met. When they are not, however, the estimates can be very inaccurate. Another limitation, shared by the capture-recapture estimates, is that the uncertainty associated with the estimated diversity increases as the proportion of unsampled taxa increases. That is, when the probability of a taxon appearing in a sample is low on average (or when there are few samples), the estimates are especially prone to error. For this reason, the fact that a method works well as an estimate of present-day diversity does not necessarily mean that it will work well when applied to the fossil record, except perhaps where that record is unusually complete.

D. Other Nonparametric Methods

While most biodiversity estimates other than the closed-population capture-recapture models are parametric, some are nonparametric. Two of these models have been applied to paleontological data. One of these, "Chao-1," was formulated by Chao (1984) and is actually mathematically equivalent to the Chao-2 estimate discussed previously. In this case, however, the number of individuals representing each taxon in a single sample is used rather than the frequency of occurrences in a set of multiple samples. That is, f_1 is the number of taxa represented by only one individual in a sample, f_2 is the number represented by two individuals, and so on. Wing and DiMichele (1994) used this approach to examine local vegetation biodiversities in the Paleozoic and Cenozoic floodplain data set discussed previously. They found that, on average, local diversity in the late Paleozoic was similar to local diversity in the early Cenozoic. This was consistent with their findings for regional diversity using the Chao-2 estimator. However, they did find greater variation in diversity levels among sites in the Cenozoic; in particular, the most species-rich Cenozoic sites were much more diverse than the most species-rich Paleozoic sites.

A second model, developed by Chao and Lee (1992), utilizes the entire frequency distribution of occurrences in a set of samples, rather than just the number occurring in only one or a few samples. Anderson and coworkers (1996) estimated diversities with this method for the late Triassic plant and insect data mentioned in the previous section. These estimates ranged from about 45% lower (for insects) to 55% higher (for plants) than estimates obtained with the GIGP model discussed previously.

These models share the limitations of the closed-population capture-recapture models discussed previously. In particular, the Chao-1 estimator assumes that probabilities of sampling are identical among samples for a particular taxon. The Chao and Lee estimator makes a similar assumption: each taxon has a constant probability of sampling associated with it. Further, it treats each occurrence as a separate sample. That is, it does not account for the fact that occurrences of taxa are grouped according to the sampling units in which they were found.

III. UNEXPLORED DIVERSITY ESTIMATES

To this author's knowledge, no other diversity estimation methods have been applied to the fossil record in published studies. However, several other methods of estimating biodiversity exist, and some may be applied to fossil data in the future. These include those capture-recapture estimates that assume all taxa have an equal probability of occurring in any given sample. When this assumption is reasonable (for instance, when studying groups of closely related taxa preserved in very similar sediments), the resulting estimates should be more precise than the methods discussed previously, which generally have greater uncertainties associated with them. Other estimates involve fitting observed distributions of occurrences to particular statistical distributions (the

GIGP method discussed previously is an example of this). These methods have been applied to estimate present-day diversity for particular taxonomic groups and particular regions, sometimes with good results. Many of these methods and applications were reviewed by Colwell and Coddington (1994) and by Bunge and Fitzpatrick (1993). The major barrier to their application to the fossil record is that differences among taxa in their probability of entering the fossil record, and subsequent loss of fossiliferous rock (at rates that may vary among regions and through time), may make distributions of fossil occurrences very different from distributions of occurrences of living organisms.

IV. CONCLUSIONS

As we have seen, many features of the fossil record make assessing diversity trends difficult. Some of these difficulties can be eased, at least in principle. For instance, unequal distribution of sampling effort by paleontologists can be reduced by emphasizing undersampled regions and strata in future fieldwork. Similarly, the expanding palette of tools for dating fossils and their surrounding sediments, their increasing precision, and the development of more robust statistical methods for chronostratigraphic correlation should progressively improve the fossil record's temporal accuracy and precision. However, some problems are less tractable. For instance, fossiliferous rock is progressively lost as it ages. Thus, the fossil records of large regions (and thus the record of biodiversity in the associated habitats) may simply not exist. This problem is particularly acute for older time intervals. Similarly, organisms in some habitats are simply less likely to be preserved than organisms in others, and those habitats will thus have poorer records of their biodiversity history.

Several approaches have been used to account for these sampling effects in the estimation of diversity or (in the case of rarefaction) to minimize the effects of variation in sample size on estimates of diversity. All of these approaches were originally designed to estimate local diversity; thus, their application to regional and global patterns is problematic, and workers have interpreted results cautiously. Nevertheless, these results can sometimes be used to eliminate some explanations for particular biodiversity trends. For instance, Raup's use of rarefaction confirmed that the increased diversity of echinoids through time was not solely due to the greater number of paleontological samples available for younger strata. When the results of these approaches differ from one another, or from diversity trends apparent from the fossil record in its entirety, however, things become complicated. Do the estimates indeed account for sampling effects and reveal true diversity trends, or does violation of model assumptions render the estimated trends even less reliable then the uncorrected diversity trends they are intended to improve?

The future of diversity estimation will no doubt involve considerable effort on several fronts. One promising approach is to apply several methods, then identify biodiversity trends that are robust to these alternative methods. Another is to investigate directly how different estimates are biased when particular assumptions are violated, then devise means of minimizing these biases. Diversity estimation methods that have not yet been applied to the fossil record may be incorporated into the paleobiological research program, and new estimates will undoubtedly be forthcoming as well. Nonparametric methods based on the frequency distribution of occurrences have been identified as a promising area for further progress by many biostatisticians. Finally, much recent work focuses on assessing the completeness of the fossil record of particular taxa or strata. In the future, these tools will undoubtedly be brought to bear on the problem of estimating diversity. For the past two decades, quantitative approaches have been rapidly growing in popularity and sophistication among paleobiologists. This movement is still in its infancy, and its future is likely to produce an increasingly clear picture of the history of biodiversity.

See Also the Following Articles

BIODIVERSITY, EVOLUTION AND • BIODIVERSITY, ORIGIN OF • EXTINCTION, RATES OF • MASS EXTINCTIONS, NOTABLE EXAMPLES OF • MEASUREMENT AND ANALYSIS OF BIODIVERSITY • PALEOECOLOGY • SPECIES-AREA RELATIONSHIPS

Bibliography

Colwell, R. K., and Coddington, J. A. (1994). Estimating terrestrial biodiversity through extrapolation. *Phil. Trans. Royal Soc. Lond. B* 345, 101–118.

Gilinsky, N. L., and Signor, P. W. (Eds.) (1991). *Analytical Paleobiology*. The Paleontological Society, Knoxville.

Guex, J. (1991). *Biochronological Correlations*. Springer-Verlag, Berlin.

Pollock, K. H., Nichols, J. D., Brownie, C., and Hines, J. E. (1990). Statistical inference for capture-recapture experiments. *Wildlife Monographs* 107, 1–97.

Raup, D. M. (1976). Species diversity in the Phanerozoic: An interpretation. *Paleobiology* 2, 289–297.

Sepkoski, J. J., Jr., Bambach, R. K., Raup, D. M., and Valentine, J. W. (1981). *Phanerozoic Marine Diversity and the Fossil Record*.

Valentine, J. W. (Ed.) (1985). *Phanerozoic Diversity Patterns: Profiles in Macroevolution*. Princeton University Press, Princeton.

FRAMEWORK FOR ASSESSMENT AND MONITORING OF BIODIVERSITY

James A. Comiskey, Francisco Dallmeier, and Alfonso Alonso
Smithsonian Institution/Monitoring and Assessment of Biodiversity Program (SI/MAB)
National Museum of Natural History

GLOSSARY

adaptive management Adaptive management is a systematic, cyclical process for continually improving management policies and practices based on lessons learned from operational programs.

biodiversity assessment A biodiversity assessment entails the identification and classification of species, habitats, and communities within a given area or region. The overall purpose is to provide information needed to evaluate whether management is necessary to conserve biological diversity. Assessments also provide data and information that can be applied to monitoring programs or for providing basic information for scientific inquiry.

biological indicators Indicators are species or communities that enable an evaluation of environmental conditions and detect changes. Indicator species are normally surrogates of other species in the area of interest and are usually sensitive to environmental change. Environmental variables may also be used as indicators.

management Management in relation to biodiversity conservation involves decisions that have consequences for biological resources. Management can be designed to protect or restore biological resources, especially under conditions where the lack of intervention would lead to an irreversible or undesired change.

monitoring Monitoring involves the repeated collection and analysis of observations and measurements to evaluate changes in populations of species and environmental conditions. Monitoring also helps in assessing progress toward meeting a management objective. Monitoring can serve as a warning system, alerting managers that changes in biodiversity may require changes in biodiversity management regimes to ensure protection of biological resources.

uncertainty Uncertainty describes the condition whereby managers have a lack of knowledge of biodiversity that prevents them from defining the best course of management action. Uncertainty can be overcome through adaptive management.

AS PRESSURES INCREASE on earth's natural resources, there is a need for managing those resources to protect biodiversity and ensure its sustainable use. A growing network of protected areas aims to ensure the survival of representative components and processes of ecosystems, especially in temperate zones. In tropical areas around the world, however, regional biodiversity is less well represented in protected areas, despite the fact that scientific records indicate biodiversity is high-

Encyclopedia of Biodiversity, Volume 3

est in the tropics. Many of these unprotected areas, important for both biodiversity and natural resource extraction, are disappearing. Well-planned monitoring and management programs will promote the long-term maintenance of biodiversity. Some such actions require the protection of native or endangered species, and the elimination or restriction of non-native invasive species.

I. INTRODUCTION

What is the most effective management approach? As you will read in this chapter, this question requires a basic understanding of ecosystem function and the participation of all interested parties from the start of the management plan. As we increase our knowledge of the ecosystem, we can make management decisions based on solid scientific knowledge. Since management and conservation of biodiversity is our ultimate goal, we must monitor the effectiveness of our conservation strategies and approaches. Thus monitoring is an integral part of biodiversity management, enabling us to learn from the outcomes of our management "experiments," a term known as adaptive management. Management practices that enhance biodiversity conservation can be continued or promoted, while those that threaten biodiversity can be evaluated and modified or discontinued.

The simplest definition of monitoring is to warn. It involves the collection and analysis of repeated measurements to evaluate the progress toward meeting a management objective. Monitoring can only be initiated if there are management objectives and alternative courses of action that can be implemented. The repeated collection of long-term data alone does not automatically imply a monitoring activity. Monitoring must be programmed within a management framework to avoid just gathering baseline information. Biodiversity changes have many causes including, among others, natural population cycles, climatic effects, and the direct impact of humans. The effects of these relationships are evaluated by research-related questions that may require multiple controls and replicates. This is an additional step in the monitoring process. This chapter examines assessment and monitoring in an adaptive management framework and identifies the key steps for a successful biodiversity monitoring project.

II. ADAPTIVE MANAGEMENT

Adaptive management is a systematic, cyclical process for continually improving management policies and practices based on lessons learned from operational activities. It is used to generate feedback for managers. The underlying assumption is that we are dealing with uncertainty and no one can be certain which management practices are best in many situations. Regular monitoring to check on the impacts of management practices will provide the information needed to determine if those practices must be changed to achieve the desired outcome. In the case of biological resources, the desired outcome is to maintain biodiversity in an optimally functioning state. Human activities can have drastic impacts on biodiversity, impacts that often are irreversible or require long and costly periods of recovery. Managing for biodiversity conservation through the adaptive management process can help avoid or mitigate those impacts.

Adaptive management can be described as a cycle of activities, where each step builds on the learning experiences of previous steps as the calibration of the goals and objectives are conducted through the monitoring feedback. The four primary steps in the adaptive management process are (a) design management and monitoring objectives, (b) implementing management, (c) assessment and monitoring, and (d) evaluation and decision making (Fig. 1). Each step is calibrated periodically to assure that the appropriate information feeds the next level. Different managers are usually responsible for each component. Thus, the cyclical nature of the process is very important in validating the results of the separate steps.

There are three types of adaptive management: reactive, passive, and active. The reactive approach is driven by factors such as public pressure and policies that are outside the management area. For example, a new law requiring protection of a certain species may lead to new management practices in a given area. Under such a scenario, the main issue may well be whether the new practices will cause conflicts among various users of the area. Passive adaptive management monitors a single

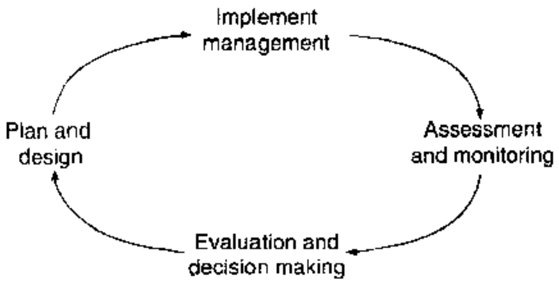


FIGURE 1 Adaptive management cycle.

management practice, which is viewed as the only alternative to achieve the management objective. Maintaining a healthy riparian community of herbaceous plants in a given area, for example, is dependent on the presence of canopy trees; thus, logging is discouraged. Water quality, which may have an effect on the herbaceous plant community, is not monitored under this approach. Therefore managers may fail to detect opportunities for improving management practices. This chapter focuses on active adaptive management, which compares alternative practices in anticipation of potential pressures from outside the management area, and taking advantage of management opportunities that may be lost under passive adaptive management.

The following section details the steps in the adaptive management process within the context of devising a biodiversity monitoring project.

A. Plan and Design

The primary task of planning and design is to identify an issue or challenge that requires implementation of a management practice to achieve a solution. The first question is what to monitor (e.g., regions, habitats, species, or guilds), followed by a determination of the scale and intensity of the monitoring activity. Establishing these parameters avoids needless investments of money, time, and other resources on collecting data that do not help evaluate whether the management is effective. This first step of the adaptive management process helps managers focus their attention on issues of concern and finding viable solutions. The following discussion describes critical elements of planning and design (Fig. 2).

1. Stakeholder Participation

Project definition, development, and implementation requires continuous stakeholder involvement and feedback. Stakeholders are individuals representing local, regional, national and international communities, agencies, and nongovernmental organizations (NGOs); industry; and any other organization with an interest in the project or responsibility within the area of the project's influence. Consulting with stakeholders throughout different stages of the project is very important. Workshops that focus on setting priorities are an excellent avenue for reviewing the various concerns of stakeholders and promoting consensus.

2. Biological Inventory and Assessment

Most traditional approaches to biodiversity assessments are oriented toward increased commercial harvesting prospects such as intense logging in temperate forests or wildlife management for waterfowl and fisheries to benefit commercial and recreational industries. More recently, biodiversity assessments have encompassed habitats and the population status of rare, endangered, and exotic or invasive species to determine which areas are important to conserve and restore. The assessment process includes literature reviews, field surveys, and inventories to gather data and information relevant to the site-specific monitoring program and to identify knowledge gaps and target indicator variables.

a. Management Plans and Literature

Priority species or habitats for monitoring may well have been identified in previous management plans or research papers. These documents should be reviewed thoroughly in the planning process to ensure that the management issues and species in question are still relevant to the current activities. Similar monitoring activities may have been conducted elsewhere and there may be valuable lessons to be learned concerning the techniques used and the responses of the species to management. At this stage, it is also possible to identify

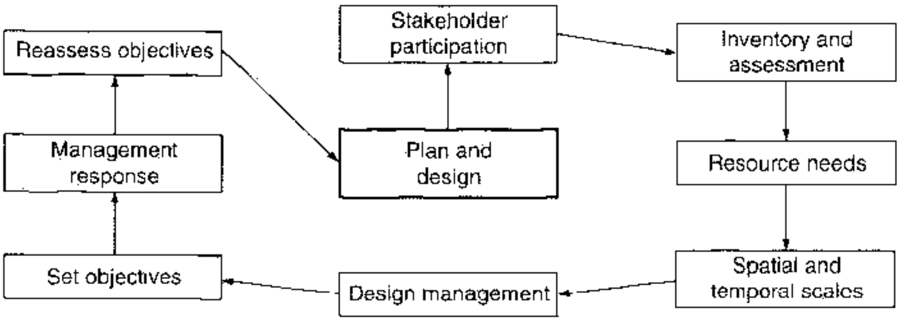


FIGURE 2 Steps involved in the plan and design phase of the adaptive management cycle.

additional stakeholders who can provide valuable advice in the development of the monitoring project. Documenting the existing information will help refine the planning process and provide a foundation for the interpretation and reporting on future results.

b. Field Assessments

Field assessments might be needed to obtain up-to-date information on the present condition of the populations or habitats in question prior to establishing specific monitoring objectives. If detailed information on species and habitats is not available, it may be necessary to conduct studies that will help identify which species are present, their abundance and frequency, and to determine which should be targeted for monitoring.

Habitat descriptions and vegetation and land-use maps can greatly enhance the process of selecting target areas for management and monitoring activities. If these are not available, the field assessment should include collection of this type of information. Sampling at a landscape scale requires this level of detail to select strata for random selection of sampling points.

c. Problem Assessment

Once available information is reviewed and any necessary field assessments are completed, planners are positioned to define the issues of management concern. The best approach often involves a workshop where specialists from various disciplines: (a) address the scope of the challenge, (b) synthesize existing knowledge, (c) identify areas of uncertainty, and (d) create a model for management. It is particularly important to develop models because they enable a clearer definition of the challenge and a more accurate evaluation of ecological understanding of the species and habitats concerned. Experts frequently help identify issues and methods for finding effective management solutions. Where necessary, strategically focused research can facilitate the development of effective monitoring designs.

d. Indicators

Sampling all components of biodiversity in a given area can be an impractical and costly task, but biodiversity surrogates enable monitoring of ecosystem functions. As examples, one might choose an oak tree species affected by gypsy moth defoliation or a species whose recruitment and mortality is impacted by extended droughts. During the planning stage, it is possible to identify potential indicator species that can be used in monitoring. Important criteria in selecting indicator species include the following:

1. The species, populations, or observed physical or chemical phenomena should be good measures of one or more questions that the monitoring program was designed to answer.
2. The indicators should be able to detect a condition in advance to assist in solving the problem or else they may have a limited role in achieving the monitoring goals.
3. It should be possible to monitor the indicators within realistic budgets.
4. Whenever possible, indicators should be selected for which there are experimental controls (e.g., populations under different management intensities).
5. The species should be selected based on their potential for impacting management decisions (charismatic species, for example, are more likely to facilitate management changes than less well-known species).

Several considerations play a role in meeting the previous criteria. First, in tropical zones where the indicator species generally are not as well known as in temperate zones, birds, large mammals, and vascular plants are frequently used as indicators for biodiversity monitoring.

Second, keystone species are those upon which many other species may rely at some point in time during their life cycles. In tropical regions, for example, nectar-feeding bats are considered keystone species because the reproductive success of many plant species depends on them through pollination. Key common species are those that are wide-ranging, easily observed and studied, long-lived, and generally occurring at high population densities (e.g., oak species in the eastern United States). The sudden increase, decline, or absence of these species in certain habitats serves as a warning that may require management attention.

Next, species or taxonomic groups that have sensitive life histories may be good indicators for biodiversity monitoring. Amphibians depend on water for reproduction and will be affected by drought or water pollution. Information on amphibian numbers, diversity, sex ratio, age, and size structure will indicate changes that may affect other components of biodiversity. And, monitoring the abundance and diversity of selected tropical frog populations can provide valuable data on the health of those populations, and hence the habitat, from one year to the next.

Finally, key habitats may also be important indicators. For example, the aspen and wet meadow habitats of the Colorado Rocky Mountains in the United States are good indicator habitats because of the species that depend on them.

3. Identifying Resource Needs

Management support of monitoring activities is essential to ensure a successful monitoring project. Essential resources include funding, people to conduct the monitoring, and technology.

Ensuring sufficient funding is crucial to long-term success. One of the most important considerations is for equipment costs, which may include not only the equipment used in the field but also computer and data-archiving facilities and associated software as well as resources for preparation and publication of the monitoring findings.

It is also important to evaluate the skills required of those who plan and implement the project, including the need for ecologists familiar with the ecosystem types under consideration, a statistician for sampling design, skilled field technicians for gathering and identifying field samples, data managers, and geographic information systems (GIS) specialists for compiling the information. Trained volunteers can be an enormous help to monitoring programs, especially in gathering data in the field, and they have been used extensively in a variety of different monitoring projects around the world.

Evaluating the time required for various phases of monitoring is appropriate during planning and design and should include time for planning, implementation, data management, and reporting. Later stages of the adaptive management cycle must receive equal attention in this regard. They frequently are not well represented in the enthusiasm of getting the project underway.

4. Defining Spatial and Temporal Scales

Biodiversity assessment and monitoring at local and regional scales can provide decision makers with high-quality data and cost-effective choices. Generally, the scale for assessment and monitoring is based on the geographical boundaries of the protected area or conservation unit or on subsets of such areas. Identifying the scale is essential at the planning stage because it has important implications in defining effective sampling design and the resources needed for the project. Monitoring species populations at a landscape scale may require sampling under different land management conditions and may call for special permits. At the local scale, monitoring usually concentrates on specific communities chosen because of the degree of threat or because of their location, and the results may not be representative of the landscape-level biodiversity.

Additional issues should be addressed when considering temporal scales or the time frame over which monitoring should be conducted. Such issues may involve not only sampling frequency but also the time taken to gather the information from each sampling point. This helps evaluate the need for qualitative (presence/absence, population estimates, or condition assessments) versus quantitative (censuses, samples, or demographics) measurements in the field.

5. Setting the Objectives

As has already been mentioned, the concept of monitoring within an adaptive management framework is very much dependent on the establishment of specific management objectives. Clearly stated objectives need to include measurable standards, desired states, threshold values, and the range and amount of acceptable change for all components of biodiversity that will be measured. The objectives should specify the limits of change before management action is taken, and they should be realistic, specific, and measurable. Several considerations assist in defining the monitoring objectives, including species and indicators to be monitored, location, attributes to be measured, action of management, the measurable quantity and degree of change, and the time frame for monitoring. These components, which should be included in the monitoring objective statement, are addressed in the following sections.

a. Location for Monitoring

Clear guidelines identifying the location of management concerns provide limits within which the management and monitoring results will be applied. These guidelines vary depending on the management responsibilities and may include a portion of the population that resides within a protected area. The location may also be determined by the purposes of the management objective and may not cover the entire range of the species within the protected area. They may, for example, be restricted to a particular area such as riparian habitats.

b. Attributes to Be Measured

For most components of biodiversity, measurable attributes comprise quantitative measures such as population abundance in the area of interest. Nevertheless, qualitative parameters may also be measured. Cover, for example, may be measured for vegetation at ground level, including the percentage cover of a given herbaceous species, or it may be measured as the proportion of forest cover determined from aerial photography.

Demography or population dynamics may be used to evaluate mortality and recruitment rates.

There are several considerations when selecting parameters to measure. The attribute should be sufficiently sensitive in detecting the desired level of change and capable of distinguishing between natural fluctuations and human induced change. The change registered by the attribute should be biologically meaningful and lead to a logical management response. The cost of measuring the variable must be within budget, and it is necessary to identify the expertise and technical ability needed to measure the attribute. In addition, the variability in observer error must be minimal. Some of these issues may be addressed by conducting pilot projects.

c. Actions and Thresholds

Populations of the selected species can increase, decrease, or remain unchanged. These parameters are important to quantify, as they will be the measures of success in a biodiversity monitoring program. Under certain conditions such as with invasive species, it may be desirable to achieve a decline in the indicator species. When the management objective is to achieve an improvement in species or habitat factors, then the increase in abundance may be the focus for management. When current conditions are favorable such as when the population of the selected species is at the desired level, the management objective should be to maintain that population density.

There may also be threshold levels beyond which the species parameters should not pass. These can be divided into biological thresholds and management thresholds. A biological threshold is the point beyond which an irreversible change in the population may occur. This may be the minimum viable population density of the species, which, if exceeded, may lead to a population crash, loss of genetic viability, or extinction. In general, biological thresholds for many species are not known. This makes it imperative to choose an indicator species with a known threshold.

A management threshold is the point beyond which management should be changed to avoid an unacceptable risk to biodiversity. Management thresholds should be maintained above biological thresholds, especially if the latter are poorly known, and management thresholds need to provide a suitable margin of error to ensure that the biological threshold is protected. The monitoring activity itself must focus on management threshold levels; for example, a population is determined to require a minimum of 100 individuals to ensure genetic diversity and prevent the local extinction of the species (biological threshold). The management threshold should be maintained above this level, say at 120 individuals, to ensure the long-term survival of the species.

d. Degree of Change

Monitoring is frequently associated with measurements, and the task of assigning a measurable range of change can be challenging because of limited knowledge of the ecology of many species. This may lead to development of objectives that are measurable and but not necessarily the most appropriate in an ecological context. Several authors have suggested that objectives do not have to be quantified, but they need to be as precisely defined as possible so that management success can be evaluated. Whatever the approach taken, the objective should consider the natural fluctuations in the population, the size necessary to ensure a viable population (thresholds), the amount of change that is biologically meaningful, the intensity of management, and the cost of measuring the specified level of change.

e. Time Frame for Monitoring

The time frame depends on the biology of the species (short-lived species will respond more quickly), the intensity of management (intense management produces rapid changes), and the level of specified change (the smaller the change, the sooner it will be detected). Short-term responses may benefit from more frequent evaluation of the management objectives, and they may be less susceptible to funding cuts. Nevertheless, sufficient time must be allocated to detect changes such as the impacts that selective clear-cutting of the forest has on water quality, which may not be detected until as much as a decade after the initial activity. Conclusions showing no problems reached after only a few years could be deceptive.

f. Statement of Management Objective

The previous steps lead to the development of a statement outlining the management objectives. An example of a management objective for monitoring could include maintaining the population of a particular species in a given area within 10% of the originally measured population density over the next 10 years. This type of objective is sometimes referred to as a target, or threshold, management objective because it measures the population in relation to a desired condition. Another management objective might be to increase the mean density of a species in a particular area by 15% over the next 5 years. This objective is referred to as a change, or trend, management objective because it measures a change in density over time. In both cases, the objective includes information on the location, the species in question, the degree of change, and a time frame for the monitoring activity.

6. Design Management

At this stage, it is important to address which options are available for management. Based on the objectives, the first decision will involve whether management should be active or passive. In most situations, active management is more appropriate as it provides several alternative options and will generally lead to quicker learning. Passive management may be suitable when it is not possible to design management with several alternatives, the costs for implementation are excessive and previous experience supports the use of one alternative as the best course of action.

The monitoring protocols to be used should be designed based on the management objectives, which address issues such as sampling design, data management, and analysis, interpretation of results, and reporting mechanisms. These issues are more fully described in the assessment and monitoring section that follows.

7. Management Response

Before monitoring begins, it is useful to identify alternative management responses based on different monitoring results. It is important for stakeholders to be aware of the implications of the monitoring results. The different management options must be evaluated to ensure that they are both economically and politically feasible. The management changes are more likely to be implemented if they are addressed in the planning stage and consensus is achieved.

8. Reassess the Objectives

It is helpful at this stage to pose a series of questions concerning the management objectives that will help identify any potential challenges or issues to be addressed. Have the objectives been clearly stated and are they realistic? What monitoring protocols are required to achieve the objectives? What is the timeline for accomplishing the objectives? Will the information that is gathered assist managers in making informed decisions? Can the results of the management decisions be statistically analyzed? Has a cost/benefit analysis been completed? What is the scale of the monitoring program (protected area, entire forest, selected number of habitats)? Is a consortium of organizations needed to achieve the objectives?

B. Management Implementation

New management should now be implemented or current management continued or modified. In most situations even if management has been in place for a number of years, the preparation and definition of management objectives will identify alternative options. Different scenarios may also be used in different locations to implement an active adaptive management approach. Where new management is identified, a detailed outline will ensure that the management activity is implemented effectively. Implementation monitoring will help guarantee the successful completion of this phase (see the next section).

C. Assessment and Monitoring

Assessment and monitoring protocols provide an essential tool for monitoring the status of biodiversity and changes in that status over time. In recent years, there has been an increased emphasis on standardizing monitoring protocols to facilitate comparisons among different projects. These long-term data are helpful in detecting the magnitude and duration of changes, how related taxa are changing, and "early warning" indicators of ecosystem health. They also serve as the basis for formulating additional research hypotheses, and, most important, the data are used to guide management decisions related to biodiversity conservation. The following section describes the different stages involved (Fig. 3).

1. Monitoring Programs

Four monitoring processes can be distinguished: baseline monitoring, implementation monitoring, effectiveness monitoring, and validation monitoring. All share the common theme of periodic measurements related to management objectives.

a. Baseline Monitoring

Baseline monitoring provides a standard against which future change may be evaluated. It is often the first set of measurements in a monitoring program and, hence, the beginning of effectiveness monitoring. Under certain conditions, baseline monitoring may be the collection of data prior to implementation of management, providing a before-and-after comparison. It may also be conducted in monitoring situations where the establishment of controls and treatments is possible.

Where a large number of variables are measured on a regular schedule with no specific management objective in mind, the activity is referred to as a long-term ecological study or a baseline study (e.g., measuring the composition, structure, and dynamics of trees in a permanent plot over a number of years). These are very important for monitoring as they provide information on a wide range of trends in biological variables.

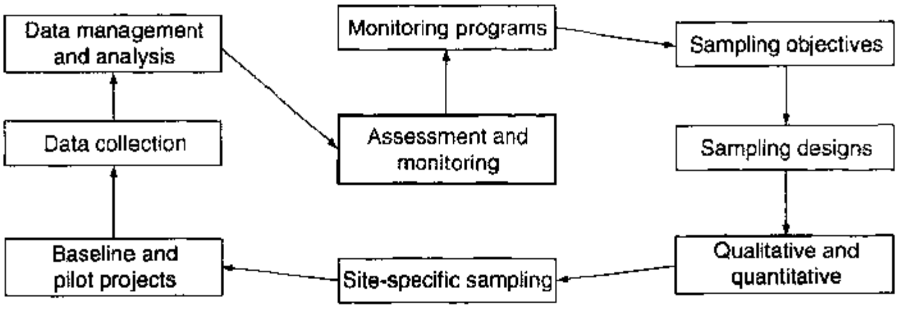


FIGURE 3 Steps involved in the assessment and monitoring phase of the adaptive management cycle.

b. Implementation Monitoring

This type of monitoring addresses the question, "Is the monitoring program doing what it is supposed to do?" It provides quantitative feedback on whether the management has been caried out as planned. Implementation monitoring acts as a form of quality control that audits the degree of compliance with previously established standards and guidelines. This type of monitoring may be conducted by external, independent parties to ensure that the implementation team is following established specifications.

c. Effectiveness Monitoring

Effectiveness monitoring assesses the outcome of the management: "Did it work?" It provides an evaluation of how effective the management was in meeting the results, needs, and expectations for which it was established. If the effects of the management do not reach the desired objective, then the management was not successful. For example, an exclusion fence for cattle or deer may be placed at the right location (implementation), but the rare plant population may still experience a decline (effectiveness). Thus, the management approach was not successful.

d. Validation Monitoring

This type of monitoring addresses the validity of predefined assumptions: "Are the models used for developing the management plan correct?" The difference between this type of monitoring and research is that research is specifically geared toward addressing cause-and-effect questions. Even though research may not fulfill the objectives of monitoring, it does provide critical information for implementing management. If monitoring requires the establishment of cause and effect, the logistical and cost implications must be evaluated carefully.

2. Defining Sampling Objectives

Sampling objectives should be defined based on the management objectives, which will specify the parameters for sampling, including the level of precision necessary. Precision is a measure of repeatability, or how close two repeated measurements are to each other, and is frequently obtained as a standard deviation of the estimated mean from which confidence limits are calculated. Increasing the number and size of samples increases confidence in the estimated measures but may be constrained by logistic and cost factors.

Sampling objectives also identify the level of change that must be detected and the risk of missing a real change or, alternatively, of detecting a false change. These factors are related to the variability of the population and the number and size of the samples used, all of which will determine the power of the sampling approach. When dealing with biological communities and attempting to maintain their parameters above the biological threshold levels, it is always best to err on the side of caution. Failing to detect a real change will result in failure of the monitoring program.

3. Sampling and Data-Gathering Designs

Sound sampling designs should provide guidelines for the most cost-efficient and effective way to gather and analyze data while maintaining high quality-control standards. The design of biodiversity monitoring programs should be based on site-specific management objectives, which help determine the sampling intensity to allow comparisons among habitat types and the assessment of long-term changes within habitats and the effects of management.

The sampling design should address the following questions:

1. What do researchers need to know about the site (e.g., data on forest composition, structure, and di-

versity as well as site-specific information such as the effects of acid deposition, drought, typhoon, El Niño, fires, and other disturbances)?
2. How will data from the monitoring program be used (managerial and scientific uses)?
3. Is the site representative of the selected habitat?
4. Will the design be sensitive enough to detect changes?
5. What limits of change are expected and are important to detect (e.g., mortality rates of forest stands should not exceed 3 to 5% per year)?
6. What is the degree of confidence expected from the results?
7. Will the sampling design produce results that will be representative of what is happening in the entire study area?

The management objectives will help determine the type of monitoring needed—qualitative or quantitative monitoring.

4. Qualitative and Quantitative Monitoring

Qualitative monitoring (e.g., habitat condition) is generally quicker and less expensive to implement than quantitative monitoring (e.g., estimates of the frequency and abundance of a species). With qualitative techniques, variability among different observers may be high. It is essential to reduce this variability because it has important implications in distinguishing real changes from those brought on by sampling variability. This may require testing several techniques to determine which one produces consistent field staff results.

Quantitative methods require more precise measurements for estimating population parameters, thus increasing the time and cost investment. Observer variability is less of an issue because direct measurements, not estimates, are recorded in the field.

5. Site-Specific Sampling

Within a selected habitat, sampling may be carried out in a systematic, random, or stratified fashion. Systematic sampling requires establishment of a grid within which a series of points are sampled. For random sampling, all points have an equal probability of being sampled. An area may also be stratified according to the needs of the monitoring program.

The number and size of the samples will be determined by the precision required for the monitoring. It will also be necessary to identify whether the samples should be permanent or temporary. Permanent sampling points will increase the level of repeatability and the cost and time needed to establish and permanently demarcate the sampling point.

Cost analysis of the sampling design is advisable so that objectives and the precision level can be adjusted, if necessary, to budget limitations. The desired precision level can be calibrated by changing the sampling plot unit and design as well as the confidence intervals and level.

6. Baseline Information and Pilot Projects

Monitoring requires a base or expected norm against which all future change may be compared. Baseline information may come from existing publications or initial surveys, assessments, or baseline monitoring. Unfortunately, initial surveys do not give any indication of how representative the results are of the selected habitat, and the literature may be sparse, especially in tropical regions. It is therefore suggested that surveillance or a pilot test be conducted before establishing a baseline against which future comparisons will be made. As an example, when establishing a vegetation plot, the first census results in a list of species and a description of how they are assembled and distributed at the site. Repeated measurements indicate changes in these variables over time and identify trends on which the baselines may be constructed.

Pilot projects are extremely useful and will likely identify issues that need to be addressed at an early stage. As a starting point, pilot projects provide guidance for the monitoring project design, the feasibility of the field techniques used, whether the sample size and number are sufficient to detect significant change, and whether the protocols used are logistically and economically feasible. Pilot projects may also identify areas where further resources need to be obtained, where training of field staff is needed, or where additional assistance is appropriate. At this stage, it is possible to modify the sampling to ensure that the management objectives can be assessed.

7. Data Collection

Data collection entails the measurement and assimilation of information in the field. A central issue in data collection is the need to gather data under consistent standards. Data collectors should be trained specifically for the task. Standard protocols should be used to facilitate cross-site comparisons and evaluation in multitaxa monitoring. For example, in past efforts related to forest management, dead and dying trees were viewed as less important than live trees because they had little value for wood products. More recent efforts related to biodiversity recognize the importance of dead trees as elements of the ecological community that support other taxa.

8. Data Management and Analysis

A primary tenet of data management is to assure that data are accurately transferred from field data sheets and securely stored, making them available for future analysis. Managing data sets from a monitoring project should include data entry, verification, validation, archiving, and documentation.

Data entry refers to the process of accurately transferring data from the field data forms to the computer. Validity checks should be incorporated to assist in the detection of errors. Data verification, which follows data entry, is the process of assuring that the computerized records match the information from the original data sheets. Data validation is performed on verified computerized data for range and logic errors. For this task, data managers must be knowledgeable about the data they are handling.

Because monitoring may take place over long periods of time, data must be safely archived. Data documentation should accompany the archive, ensuring that all the information required for understanding the data is stored for future reference. Data documentation incorporates a number of descriptors such as the data set name, title, files and their structure, location, investigator, and dates of monitoring along with any other information pertinent to the data. A data dictionary describing the structure of the database should be stored with the archived data. When planning a monitoring program, researchers must include the costs involved in data management in long-term budgets. It is money well invested.

Data analysis is essential in relating data to the established goals and objectives of the monitoring program. Results from data analysis, recorded in the form of syntheses or reports, need to be available to other researchers and to managers and decision makers. Data analysis should: (a) ensure that the monitoring objectives are effectively addressed, (b) emphasize the importance of early detection of critical changes to allow lower-cost solutions, (c) recognize that results from the analysis are meant to contribute to knowledge about the biodiversity component being monitored, and (d) emphasize that the results are intended to provide biodiversity managers with a scientific rationale for setting appropriate standards.

While data analysis should synthesize large quantities of information to make it meaningful to the user, sometimes the tendency is to over analyze it. This often produces an overwhelming amount of analytical results and interpretations that can obscure meaningful findings. It is important to maintain the focus of the monitoring program by returning to the specific questions that need to be answered. The results from the data analysis should be presented in a format that enables others to make their own interpretations without difficulty. This may involve the preparation of reports detailing how the findings affect the management of the protected area and presenting recommendations.

D. Evaluation and Decision Making

Evaluation elicits answers to the questions underlying the project's objectives, thus allowing for generation of management recommendations and calibration of the monitoring program. In addition, it allows an assessment of the monitoring process. It is often useful to pose the following questions: Are the results of each specific monitoring technique well integrated with the overall monitoring program? Do methods used ensure reliable, timely, and effective data analysis? Are the collected data subject to the appropriate techniques for data management and analysis? Could the data gathered be coupled with new technologies for analysis and management? What mechanisms exist or can be developed to allow timely transfer of data and information to managers and decision makers?

Management approaches should be viewed as hypothetical means to reach operational goals. Through evaluation of the monitoring data, managers receive timely feedback as these hypothesis are tested. Thus, evaluations are the tools for improving management by checking on management actions and providing guidelines for improvement. For example, when a predetermined degree of change is detected, appropriate action is taken and the results must be evaluated.

All the preceding steps lead to decision making regarding the need or lack thereof to adjust the management practices and monitoring program. If the findings determine that biodiversity trends are within the expected values, monitoring will continue without substantial alterations. If significant changes in the trends are observed, managers or decision makers need to design the most appropriate response. The reasons for monitoring can be evaluated at this stage. Is the monitoring still required, and, if so, do the objectives still remain the same?

Inconclusive results require adjustment of the objectives and sampling approaches to increase the degree of precision. However, careful planning and design of the monitoring program can reduce the risk of inconclusive information.

III. KEYS TO SUCCESS

One of the most critical and challenging issues in the success of monitoring programs is the need to gain long-term support. Project personnel must elicit the endorsement of high-level managers and the commitment of resource experts and technicians in the field. One method of achieving this goal is to incorporate monitoring as a routine duty of appropriate personnel. Another is to implement a reward system for personnel who detect a situation that needs to be fixed and take steps to devise a management-oriented solution. Institutional commitment is essential to ensure the long-term success of any monitoring activity.

A common practice is to develop partnerships among several organizations to assist in fund raising, standardizing protocols, providing site-specific documentation, and supporting training for high-quality research and monitoring. This encourages sound data and information management and reduced duplication. Such partnerships have also aided in disseminating information about a monitoring program to inform the public of its importance and benefits. Greater public understanding may lead to increased acceptance and support for the program.

Creating an interdisciplinary team with an integrated vision of the overall project beyond individual interests greatly facilitates the monitoring success. This type of integration is frequently achieved through workshops or through training of staff and managers in leadership and persuasive communications skills that promote teamwork in dealing with conflict resolution that is common to interdisciplinary teams.

It is also clear from the framework outlined here that planning is important to monitoring success. Revisiting the management and monitoring objectives will ensure that the activity proceeds smoothly. By using the principles of adaptive management, the results of the monitoring will help determine the success of the management and refine future activities to ensure the conservation of biological diversity.

Acknowledgments

We thank L. Alonso, P. Campbell, G. Davis, D. C. MacIver, C. Drysdale, and D. Kloepfer for providing valuable comments on the manuscript. We also thank numerous other people who have helped in advancing these concepts over the years.

See Also the Following Articles

ECOSYSTEM, CONCEPT OF • ECOSYSTEM FUNCTION MEASUREMENT, AQUATIC AND MARINE COMMUNITIES • ECOSYSTEM FUNCTION MEASUREMENT, PLANT COMMUNITIES • ECOSYSTEM FUNCTION, PRINCIPLES OF • POPULATION VIABILITY ANALYSIS (PVA)

Bibliography

Alonso, A., and Dallmeier, F. (2000). *Working for Biodiversity*. Smithsonian Institution/Monitoring and Assessment of Biodiversity Program, Washington, D.C.

Dallmeier, F., and Comiskey, J. A. (Eds.) (1998). Forest biodiversity research, monitoring and modeling: Conceptual background and old world case studies. *Man and the Biosphere Series*, Vol. 20. UNESCO and The Parthenon Publishing Group, Carnforth, Lancashire.

Elzinga, C. L., Salzer, D. W., and Willoughby, J. H. (1998). *Measuring and Monitoring Plant Populations*. BLM Technical Reference 1730-1. Bureau of Land Management, Denver, CO.

Goldsmith, F. B. (Ed.) (1991). Monitoring for conservation and ecology. *Conservation Biology Series* 3. Chapman & Hall, London.

Heyer, W. R., Donnelly, M. A., McDiarmid, R. Y., Hayek, L. C., and Foster, M. S. (1994). Measuring and monitoring biological diversity: Standard methods for amphibians. In *Biological Diversity Handbook Series* (M. S. Foster, Ed.). Smithsonian Institution Press, Washington, D.C.

Holling, C. S. (Ed.) (1978). *Adaptive Environmental Assessment and Management*. John Wiley & Sons, New York.

Johnson, N., Malk, A., Sexton, W., and Szaro, R. (1999). *Ecological Stewardship: A Common Reference for Ecosystem Management*, Volume I. Elsevier Science, Oxford.

Noss, R. F., and Cooperrider, A. Y. (1994). *Saving Nature's Legacy: Protecting and Restoring Biodiversity*. Island Press, Washington, D.C.

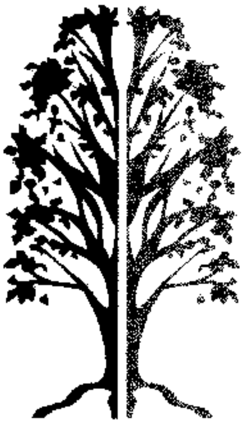
Sexton, W. T., Malk, A. J., Szaro, R. C., and Johnson, N. C. (1999). *Ecological Stewardship: A Common Reference for Ecosystem Management*, Volume III. Elsevier Science, Oxford.

Spellerberg, I. F. (1992). Evaluation and assessment for conservation. *Conservation Biology Series* 4. Chapman & Hall, London.

Szaro, R. B., and Johnston, D. W. (Eds.) (1996). *Biodiversity in Managed Landscapes: Theory and Practice*. Oxford University Press, New York.

Walters, C. J. (1986). *Adaptive Management of Renewable Resources*. McGraw-Hill, New York.

Wilson, D. E., Cole, F. R., Nichols, J. D., Rudran, R., and Foster, M. S. (1996). Measuring and monitoring biological diversity: Standard methods for mammals. In *Biological Diversity Handbook Series* (M. S. Foster, Ed.). Smithsonian Institution Press, Washington, D.C.



FRESHWATER ECOSYSTEMS

Robert G. Wetzel
University of Alabama

GLOSSARY

benthos Nonplanktonic animals associated with substrata within sediments or closely above the sediment–water interface.

detritus Nonliving organic matter in both soluble and particulate forms.

littoral zone Region of a lake or river between the land and the open water (pelagic zone) that is colonized by emergent, floating-leaved, and submersed aquatic plants and their attendant sessile microbiota (periphyton).

pelagic zone Open-water portion of a lake or reservoir beyond the littoral zone.

periphyton Bacteria, fungi, algae, and sessile microfauna growing attached to substrata (sediments, rock, plants, animals, sand).

phytoplankton Small photosynthetic plankton, largely algae and cyanobacteria.

plankton Small organisms with no or limited powers of locomotion that are suspended in the water and largely dispersed by turbulence and other water movements.

production Amount of new organic biomass formed over a period of time, and includes any losses from respiration, excretion, secretion, injury, death, and predation.

zooplankton Small animal plankton, usually denser than water, that sink by gravity to greater depths.

FRESHWATER ECOSYSTEMS ARE INTERACTIVE SYSTEMS within which biotic species and their growth and adaptation, and associated biological productivity, nutrient cycling, and energy flows among inland aquatic microbial, plant, and animal communities, are integrated with their environment. These inland waters include lakes, reservoirs, rivers, streams, and wetlands. The subject discipline, termed *limnology*, evaluates how these interactive relationships are regulated by natural and modified variations in their physical, chemical, and biotic environments. The standing water (*lentic*) lake or reservoir ecosystem is a system that is intimately coupled with the land surrounding it in its drainage area via chemical components transported to the lake by groundwater or running surface waters (*lotic*) of streams and rivers.

I. FRESH WATERS: PHYSICAL ECOSYSTEM STRUCTURE

Most water on the Earth's surface is saline in oceans (97.61%), and most of the remainder is polar ice

(2.08%) and ground water (0.29%). The remaining water (only 0.009%) exists in freshwater ecosystems as temporary storage in lakes and reservoirs on land. These surface waters have much shorter renewal times than those of the oceans (Table I). The volume of water flowing from these storage sites to the sea in streams and rivers is very small (0.00009% of total water), with a mean residence time of about two weeks. During high-precipitation periods, surface waters of rivers increase and often recharge adjacent groundwater aquifers. Return flows from ground water to rivers occur during periods of low flow or drought, and usually maintain a base flow in river channels.

The storage of water in lakes and retention times are altered by shifts in the balance between inputs from all sources and water losses. Lakes receive water from precipitation directly on the surface, from surface inflows from the drainage basin, and from subsurface groundwater seepage. Lakes lose water by flow from a usually single outlet (*drainage lakes*), by seepage through the basin walls into the ground water (*seepage lakes*), evaporation, and evapotranspiration from higher aquatic plants. Saline lakes, which constitute nearly half of inland surface waters (see Table I), occur in *closed* basins with no outflow except by evaporation. That evaporation results in a gradual marked increase in residual concentration of dissolved salts that were imported with influent water.

Natural lakes are concentrated in the subarctic and temperate regions of the Northern Hemisphere, whereas reservoirs are constructed predominately in the subtemperate and subtropical regions. Some 40% of the total volume of surface freshwater is contained in seven great lake basins of Siberia (Lake Baikal), North America (the Laurentian Great Lakes: Superior, Michigan, Huron, Erie, and Ontario), and eastern Africa (Lake Tanganyika). Most of the millions of lakes and reservoirs, however, are very much smaller and relatively shallow, usually <15 m in depth (Fig. 1). Mean depths are even more shallow (<5 m). As a result, light commonly penetrates to over half of the sediments within the basins of a large percentage of lakes. Photosynthesis can thus occur not only by algae that are suspended in the water column but can extend to higher aquatic plants and attached algae. Major flowing-water ecosystems occur in low-gradient tropical and subtropical regions and are posited to support high photosynthetic productivity in extensive floodplain and land–water interface regions.

The productivity and internal metabolism of aquatic ecosystems are driven and controlled by energy from solar radiation acquired by photosynthesis. Inland wa-

TABLE I

Water in the Biosphere[a]

	Volume (thousands of km^3)	Percentage of total	Renewal time
Oceans	1,370,000	97.61	3100 years[b]
Polar ice, glaciers	29,000	2.08	16,000 years
Groundwater (actively exchanged)[c]	4000	0.295	300 years
Freshwater lakes	125	0.009	1–100 years[d]
Saline lakes	104	0.008	10–1000 years[d]
Soil and subsoil moisture	67	0.005	280 days
Rivers	1.2	0.00009	12–20 days[e]
Atmospheric water vapor	14	0.0009	9 days

[a] Modified from Wetzel (1983, 2000) after Kalinin and Bykov. Slightly different values are given by I. A. Shiklomanov, but ratios are similar.

[b] Based on net evaporation from the oceans.

[c] Kalinin and Bykov estimated that the total groundwater to a depth of 5 km in the earth's crust amounts to 60×10^6 km^3. This is much greater than the estimate by the U.S. Geological Survey of 8.3×10^6 km^3 to a depth of 4 km. Only the volume of the upper, actively exchanged groundwater is included here.

[d] Renewal times for lakes vary directly with volume and mean depth, and inversely with rate of discharge. The absolute range for saline lakes is from days to thousands of years.

[e] Twelve days for rivers with relatively small catchment areas of less than 100,000 km^2; 20 days for major rivers that drain directly to the sea

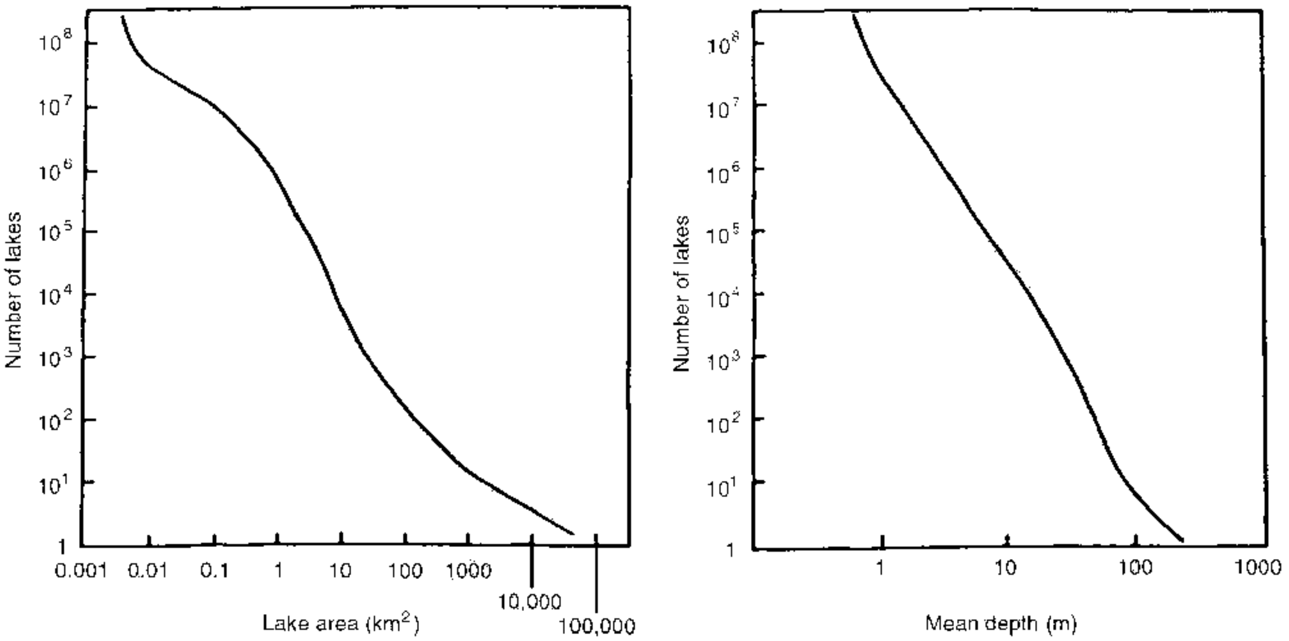


FIGURE 1 Approximate number of lakes of the world in relation to lake area and approximate mean water depth. (After Wetzel, 1990).

ters receive organic products of photosynthesis directly from their aquatic flora and indirectly from their drainage basins as particulate and dissolved organic matter from terrestrial and wetland plants imported by stream water, storm runoff, ground water, and the atmosphere.

Light is attenuated exponentially with increasing depth in water. As the light is attenuated, ultraviolet wavelengths are absorbed strongly by dissolved organic compounds in the water. The blue portion (500–600 nm) of the visible spectrum penetrates most deeply in relatively clear waters. Infrared wavelengths are absorbed rapidly, largely as a result of the molecular structure of water molecules, and much of this energy is dissipated as heat. Because water becomes less dense as heat content increases above 4°C, the less dense warmer water floats upon the more dense cooler water. This density stratification results in separation of major strata within lakes, where a warmer, less dense stratum (*epilimnion*) overlies a cooler, denser zone (*hypolimnion*). The interface zone (*metalimnion*) is a region of rapid thermal discontinuity, often as much as several degrees change per meter, between the epilimnion and the hypolimnion.

The resulting density stratification affects not only the thermal structure and water mass stratification, but also the hydrodynamics of lakes and reservoirs. Heat is distributed and altered by the physical work of wind energy, currents and other water movements, basin morphometry, and water losses. Patterns of density-induced stratification influence physical and chemical properties and cycles both spatially within the lake and seasonally. These characteristics structure the aquatic habitats and have marked attendant effects on all chemical cycles, metabolic rates, and the population dynamics of organisms and their productivities.

The total *salinity* of inland waters nearly always consists of eight most abundant ionic species, of usually four major cations (Ca^{2+}, Mg^{2+}, Na^+, K^+) and four anions (HCO_3^-, CO_3^{2-}, SO_4^{2-}, Cl^-). The salinity of fresh waters has a world average concentration of about 120 mg/L, but varies among continents and with the lithology of landmasses (Table II).

Of the major constituents of the cellular protoplasm of organisms (C, H, O, N, P, and S), the biogeochemical cycles control the physical and metabolic availability of phosphorus, nitrogen, and several minor nutrients. These elements often limit biotic development, particularly photosynthetic generation of organic matter. Phosphorus, in comparison to other macronutrients required by biota, is least abundant geochemically in a large portion of the global landforms and drainage. As a result, phosphorus is commonly the first element to limit biological productivity, especially of photosynthetic producers and heterotrophic microbiota. When phosphorus is in adequate supply, nitrogen availability invariably limits productivity. Standing inland waters

TABLE II

Mean Composition of River Waters of the World (mg L^{-1})[a]

	Ca^{2+}	Mg^{2+}	Na^+	K^+	CO_3^{2-} (HCO_3^-)	SO_4^{2-}	Cl	NO_3	Fe (as Fe_2O_3)	SiO_2	Total
North America	21.0	5	9	1.4	68	20	8	1	0.16	9	142
South America	7.2	1.5	4	2	31	4.8	4.9	0.7	1.4	11.9	69
Europe	31.1	5.6	5.4	1.7	95	24	6.9	3.7	0.8	7.5	182
Asia	18.4	5.6	5.5	3.8	79	8.4	8.7	0.7	0.01	11.7	142
Africa	12.5	3.8	11	—	43	13.5	12.1	0.8	1.3	23.2	121
Australia	3.9[b]	2.7	2.9	1.4	31.6	2.6	10	0.05	0.3	3.9	59
World	15	4.1	6.3	2.3	58.4	11.2	7.8	1	0.67	13.1	120
Cations (μeq L^{-1})	750	342	274	59	—	—	—	—	—	—	1425
Anions	—	—	—	—	958	233	220	17	—	—	1428

[a] From Wetzel (1983, 2000), after data from several sources.

[b] Values of calcium are likely less, on the average, than Na and Mg in Australian surface waters.

have been categorized into various trophic scales on the basis of ranges of major nutrients and algal productivity (Table III).

Functionally similar organisms of biological communities can be grouped into *trophic levels* based on similarities in patterns of organic matter production and consumption. Energy is transferred and nutrients are cycled within an overall ecosystem trophic structure. The productivity of each trophic level is the rate at which energy enters the trophic level from the next lower level. Because organisms expend considerable energy for maintenance and since death of an organism routes much energy and nutrients into the detrital pool of dead organic matter, only a portion of the energy of one trophic level is available for transfer and use by higher trophic levels. Available energy decreases progressively at higher trophic levels, so that rarely can more than five or six trophic levels be supported. The efficiency of energy transfer from one level to the next is low (5–15%), and often decreases as trophic level increases.

The organic matter of freshwater ecosystems is primarily (>90%) dead organic matter, and most (usually 80–90%) of this organic detritus occurs as soluble or colloidal organic compounds. Oxygen, however, is very insoluble in water, and most diffuses to the atmosphere. When organic matter in lakes is metabolized and organic compounds respired by heterotrophic organisms, the amount of organic matter to be oxidized can be much larger than the dissolved oxygen available for oxidation. This limitation can become acute when lakes are stratified by density differences, and exposure of deep strata to oxygen of the atmosphere may be restricted for long periods of time (months).

In large lakes, the volume of water with sufficient light to support photosynthesis (*euphotic zone*) is small relative to the total volume of water containing stored dissolved oxygen that can be used in heterotrophic respiration. In small and relatively shallow lakes, however, the proportion of water supporting photosynthesis is large relative to total lake volume, and respiratory demands can exhaust oxygen dissolved in the lower strata of the lake. The process of eutrophication exacerbates the exhaustion of dissolved oxygen in large portions of lake ecosystems. *Eutrophication* is the increased loading of organic matter to an aquatic ecosystem. Often in lakes, eutrophication results from increased photosynthetic production of organic matter in response to excessive loading of nutrients, particularly phosphorus, to the water from external sources.

When dissolved oxygen is depleted from water strata, alternate electron acceptors are utilized by microbial metabolism of dissolved organic compounds. Bacterial denitrification occurs by biochemical reduction of oxidized nitrogen anions (NO_3 and NO_2), concomitant with the oxidation of organic matter by many genera of facultative anaerobic bacteria. When nitrate and nitrite are depleted, alternative oxidized ions can function as electron acceptors, although with decreasing efficiencies—manganese, iron, and particularly sulfate, the latter being reduced to hydrogen sulfide. Often the same compound can serve as a hydrogen acceptor or donor,

TABLE III

General Ranges of Primary Productivity of Phytoplankton and Related Characteristics of Lakes of Different Trophic Categories[a]

Trophic type	Mean primary productivity (mg C m^{-2} day^{-1})[b]	Phytoplankton density (cm^3 m^{-3})	Phytoplankton biomass (mg C m^{-3})	Chlorophyll[a] (mg m^{-3})	Dominant phytoplankton	Light extinction coefficients (η m^{-1})	Total organic carbon (mg L^{-1})	Total P (μg L^{-1})	Total N (μg L^{-1})	Total inorganic solids (mg L^{-1})
Ultraoligotrophic	<50	<1	<50	0.01–0.5		0.03–0.8		<1–5	<1–250	2–15
Oligotrophic	50–300		20–100	0.3–3	Chrysophyceae, Cryptophyceae,	0.05–1.0	<1–3			
Oligomesotrophic		1–3			Dinophyceae, Bacillariophyceae			5–10	250–600	10–200
Mesotrophic	250–1000		100–300	2–15		0.1–2.0	<1–5			
Mesoeutrophic		3–5						10–30	500–1100	100–500
Eutrophic	>1000		>300	10–500	Bacillariophyceae, Cyanobacteria,	0.5–4.0	5–30			
Hypereutrophic		>10			Chlorophyceae, Euglenophyceae			30–>5000	500–>15,000	400–60,000
Dystrophic	<50–500		<50–200	0.1–10		1.0–4.0	3–30	<1–10	<1–500	5–200

[a] From Wetzel (1983, 2000), after many authors and sources.

[b] Referring to approximately net primary productivity, such as measured by the ^{14}C method.

depending on the environmental conditions. When these electron acceptors are essentially depleted, organic matter can be anaerobically degraded by fermentation to methane and CO_2 in two stages. Facultative and obligate anaerobic bacteria convert proteins, carbohydrates, and fats primarily to fatty acids by hydrolysis and fermentation. Obligatory anaerobic methanogenic bacteria then convert the organic acids to methane and CO_2.

II. DIVERSITY OF HABITATS

Freshwater ecosystems consist of entire drainage basins as water moves from land and in groundwater runoff to stream and river channels, and to recipient lakes or reservoirs. The nutrient and organic matter content of drainage water from the catchment area is modified in each of the terrestrial soil, stream, and wetland–littoral components as water moves downgradient to and within the lake or reservoir itself (Fig. 2). Photosynthetic productivity of organic matter is generally low to intermediate in the terrestrial components, highest in the wetland–littoral interface regions between the land and water, and lowest in the open water (*pelagic*) zone. The same productivity profile emerges in the gradient from land to river channels, where the greatest productivity occurs in the marginal floodplain regions. Autotrophic productivity in river channels is generally low, as is also the case in the pelagic regions of lakes. Most of the organic matter utilized by heterotrophic communities in running water is imported from floodplain and terrestrial sources as particulate and especially dissolved and colloidal organic compounds.

The interface region between the land and water is always the most productive per unit area along the gradient from land to the open water of lakes, reservoirs, and streams. Because most aquatic ecosystems occur in geomorphologically mature terrain of gentle slopes and are small and shallow, the wetland–littoral components usually dominate in both productivity and the synthesis and loading of organic matter to the systems. The region of greatest productivity is the emergent macrophyte zone. Emergent aquatic plants have a number of structural and physiological adaptations that not only tolerate the hostile reducing anaerobic conditions of saturated sediments but also exploit the high nutrient conditions and water availability of this habitat. Nutrients entering the zone of emergent macrophytes from external sources tend to be assimilated by bacterial and algal microflora of the sediments and detrital organic particles, and are then recycled to the emergent macrophytes. Dissolved organic compounds released from decomposition of plant detrital materials dominate the export of organic matter from the emergent plant zone.

Submersed macrophytes are limited physiologically by slow rates of diffusion of gases and nutrients in water within the boundary layers surrounding the leaves and by reduced availability of light underwater. Internal recycling of resources, particularly of gases (CO_2, O_2) of metabolism and of critical nutrients, is important to the abilities of submersed plants to function and grow as well as they do in underwater conditions of chronic light and gas limitations. Despite these adaptive mechanisms, growth and productivity of submersed plants are less than those of emergent and floating macrophytes.

The second most productive component of the wetland–littoral community is the microflora attached to aquatic plants epiphytically and to other surfaces, both living and dead. The surfaces provided by aquatic plants in lakes and rivers can be very large, often exceeding 25 m^2 per square meter of bottom sediments. High sustained growth of attached microflora results from their recycling of essential gases (CO_2, O_2) and dissolved nutrients within the attached communities. Nutrient uptake from the surrounding water is directed primarily to the high net growth of attached microflora and is responsible for the high capacity of wetland–littoral areas to improve the quality of water passing through these communities.

The wetland–littoral complex of higher plant and microbial communities produces the major sources of organic matter and energy of many freshwater ecosystems, including the marginal floodplains of many rivers. Most of the particulate organic matter is decomposed within these interface regions. Organic matter is exported predominantly from these marginal regions as dissolved organic matter to the recipient lake or river (Fig. 3).

The deep-water pelagic zone of lakes is the least productive along the gradient from land to water (see Fig. 2), regardless of nutrient availability. Growth of phytoplanktonic algae of the pelagic zone is limited by sparse distribution in a dilute environment where efficient nutrient recycling is restricted by the sinking of senescent phytoplankton below the depth of photosynthesis. When nutrient recycling and availability are increased, greater phytoplankton cell densities attenuate underwater light and reduce the volume of water in which photosynthesis occurs. Despite low productivity per unit area, pelagic productivity can be collectively important in large lakes and for higher trophic levels that depend on this source of organic matter.

Higher trophic levels of communities in freshwater

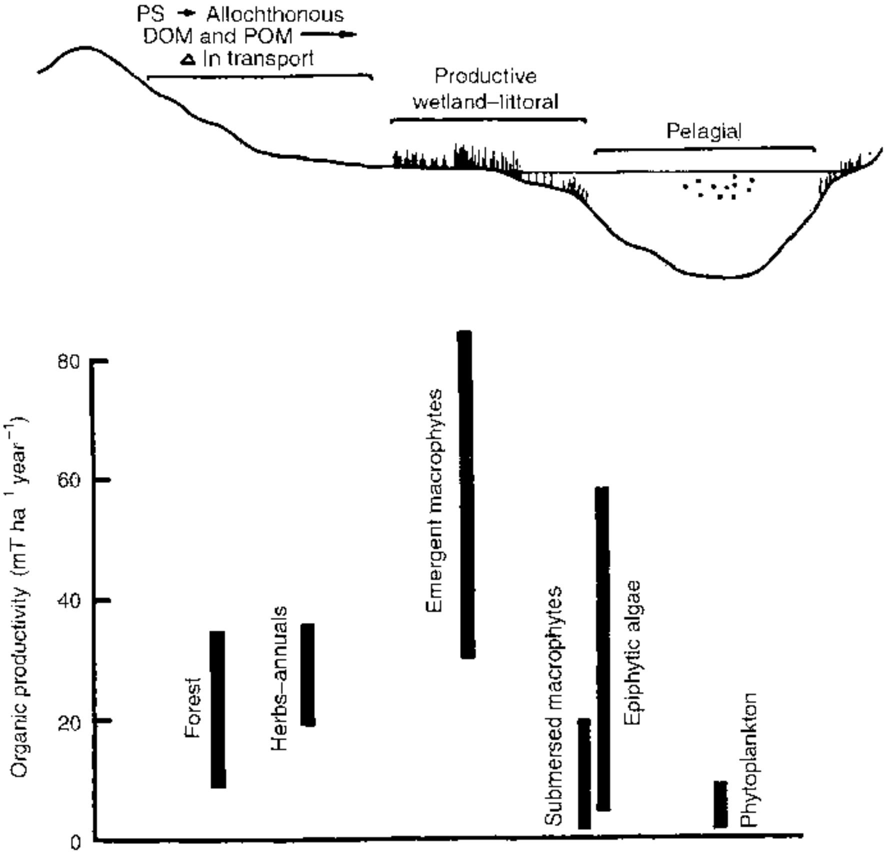


FIGURE 2 The lake ecosystem showing the drainage basin with terrestrial photosynthesis (PS) of organic matter, movement of nutrients and dissolved (DOM) and particulate (POM) organic matter in surface water and groundwater flows toward the lake basin, and chemical and biotic alteration of these materials en route, particularly as they pass through the highly productive and metabolically active wetland–littoral zone of the lake per se (net organic productivity in metric tons per hectare per year). (Modified from Wetzel, 1990).

ecosystems consist of zooplankton (dominated by four major groups of animals: protozoa/protista, rotifers, and the crustaceans cladocera and copepoda) and benthic invertebrates. In the pelagic zone, small fishes, fry of larger fishes, and predatory zooplankton, which collectively comprise a third trophic level (primary carnivores), consume a portion of these generally herbivorous organisms. A fourth trophic level may consist of medium-sized piscivorous fishes, and the fifth level of large predatory piscivorous fishes. Higher trophic levels are rare in freshwater ecosystems.

The species composition of the higher trophic levels affects the pathways of energy utilization from lower trophic levels. Environmental factors that selectively influence the populations of the communities can alter the pathways and strengths of energy fluxes from subordinate trophic levels. For example, efficiency of consumption of primary production by zooplankton is often appreciably greater in the absence of zooplankton-feeding fishes than in their presence. The population structure of the phytoplankton community responds variably to grazing impacts in concert with their available resources (light, nutrients, and organic constituents). The phytoplankton community may or may not be able to compensate for grazing losses in overall primary production, but generally is able to shift quite quickly to an alternative, less vulnerable species composition.

III. ECOSYSTEM FUNCTIONS: INTERACTIVE REGULATORY MECHANISMS

Most of the organic carbon of aquatic ecosystems, and hence much of the energy potentially available for me-

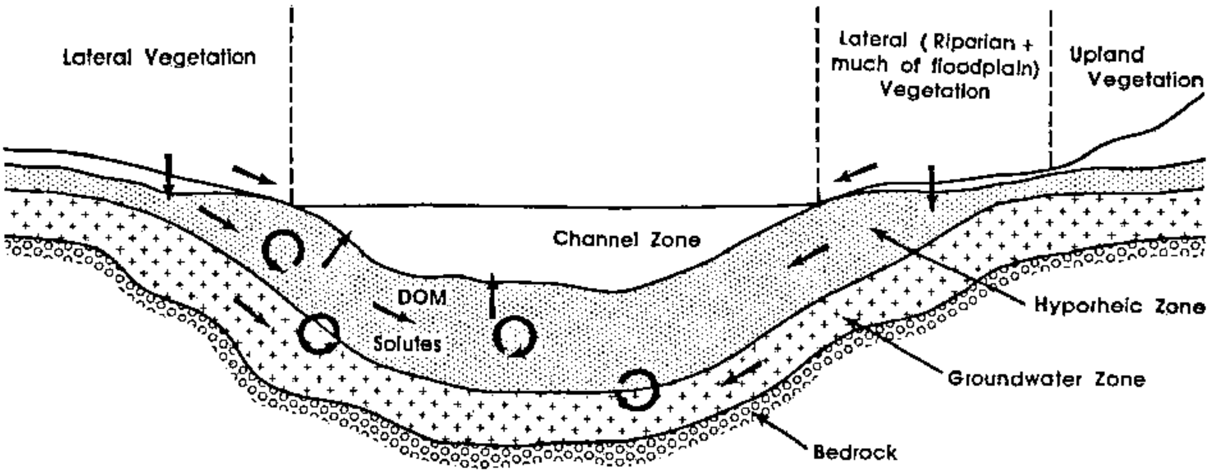


FIGURE 3 Lateral and vertical boundaries of flowing-water ecosystems. The stream ecosystem boundary is defined as the hyporheic/goundwater interface and thereby includes a substantial volume beneath and lateral to the main channel. Vegetation rooted in the hyporheic zone is therefore part of stream ecosystem production. Arrows indicate flow pathways of dissolved organic matter and inorganic solutes derived from plant detritus within the stream ecosystem. [From R. G. Wetzel and A. K. Ward (1992). In *Rivers Handbook*, *I* (P. Calow and G. E. Petts, eds.), pp. 354–369. Blackwell Scientific, Oxford, England].

tabolism of heterotrophic biota of the ecosystem, exists as dead organic matter (*detritus*), in both dissolved and particulate forms. Dissolved organic matter and much of the particulate organic matter, usually expressed as organic carbon, move with the water. Much of the particulate organic matter that is not decomposed ultimately settles by gravity and is deposited at the bottom of static water bodies. Dissolved organic compounds can also sediment if adsorbed to inorganic or organic particulate matter or if polymerization and aggregation occur into particulate form. Because much of the highest photosynthetic production of organic matter occurs in shallow regions of lakes and streams, and because most lakes are small and shallow, most heterotrophic decomposition occurs in benthic regions as organic matter. Therefore, because much of the organic matter is displaced to the sediments, most of the total heterotrophic decomposition is also largely displaced from sites of production to the sediments.

The specific composition of organic matter varies greatly. The organic matter of pelagic phytoplankton, which can dominate in-lake production in very large, deep lakes, is relatively labile. Much of the phytoplanktonic algal and cyanobacterial organic matter is decomposed by microbiota either in the water column within the water strata of production or below as it sediments to the bottom. Some highly variable portion, usually much less than 50% on an annual basis, of the phytoplanktonic organic matter is ingested and partially digested by zooplankton. An appreciable portion, usually >50%, of that particulate organic matter ingested by zooplankton is egested as dissolved and as particulate organic matter, some of which also settles as feces to the sediments.

The composition of organic matter of the wetland and littoral regions of lakes and streams tends to be considerably more recalcitrant to rapid microbial degradation than that of phytoplankton because of the littoral dominance of lignocellulose-based organic compounds of structural tissues of higher plants. Because of the much higher rates of production and slower rates of decomposition of both the particulate and dissolved organic matter from these shallow-water sources, and because of the shorter distances and times for particulate organic matter to be displaced to the sediments, organic matter accumulates on and in the sediments at greater rates in shallow regions than in those of the pelagic region. The deposition of organic matter in the sediments encourages intensive bacterial metabolism. As a result, intrusion of dissolved oxygen from the overlying water is consumed rapidly; nearly all sediments in standing freshwater ecosystems are anaerobic and highly reducing chemically.

Although the specific composition of organic matter varies greatly, detritus inevitably carries most of the energy of the ecosystem from its points of photosynthetic origins to places of transformation by heterotrophic organisms. Most of that heterotrophic transformation occurs by microbes in areas where organic matter is concentrated. Even though greater than 70% of the very large pool of dissolved organic matter of the pelagic region of lakes and rivers consists of relatively recalcitrant humic and fulvic acids that originated from higher plants, these compounds are slowly metabolized by bacteria, often being catalyzed by partial photolytic degradation by ultraviolet light (both UV-A and UV-B). In contrast, much of the particulate organic matter of the productive littoral areas accumulates on the bottom, largely under anaerobic conditions particularly in sediments and in detrital mats.

Transformation of particulate organic matter of the pelagic zone also occurs by digestion in predatory metazoan animals as well as by pelagic microbes. Productivity of the microbes is maximized within the resources available from dead organic matter and nutrients. Microbial heterotrophic utilization is highly dynamic and changes with sufficient rapidity to allow many generation turnovers before predators with much slower turnover rates can respond reproductively to changes in availability. Viral mortality of microbes, as well as ingestion of bacteria by protists (the "microbial loop"), diverts much organic matter from animal trophic levels. Bacterial, viral, and much of protistan heterotrophic metabolism as respiration represents a major output of carbon, largely as CO_2, from the aquatic ecosystem. Although this output is a loss of carbon from the ecosystem, it is not an energetic loss from the ecosystem.

That productivity that is not ingested by metazoan animals is of major importance and dominates the energy flows within the pelagic region. Nonpredatory death and metabolism by prokaryotic and protistan heterotrophs usually dominate energy flows within the pelagic region and completely dominate energy flows of the composite wetland–littoral, benthic, and pelagic components that collectively constitute freshwater ecosystems. Even ignoring the benthic portions of lakes where most of the organic carbon metabolism occurs and considering only the pelagic metabolism of lakes, for example, most of the organic carbon entering the system does not reach higher trophic levels.

The microbial heterotrophy is of primary importance to higher trophic levels by means of feedback processes, both positively (e.g., nutrient recycling and utilization by primary producers) and negatively (e.g., oxygen consumption and production of toxic fermentative metabolic end products). The abundance, distribution, and microbial decomposition of dissolved and particulate detritus both regulate and stabilize energy metabolism and nutrient availability in aquatic ecosystems. Because of the very large magnitudes and relative chemical recalcitrance of these detrital sources, the large but slow metabolism of detritus provides an inherent ecosystem stability that energetically dampens the ephemeral, volatile fluctuations of numbers and biomass of the biota of higher trophic levels.

IV. BIODIVERSITY WITHIN FRESHWATER ECOSYSTEMS

The biodiversity of most microbial, plant, and animal groups of freshwater stream, lake, and wetland ecosystems is very poorly known on a global basis. Moreover, it is likely erroneous to believe that biodiversity of many taxa in temperate ecosystems is appreciably less than that in tropical freshwater ecosystems. For example, the rivers and streams of Alabama, a "hot spot" within the United States, contain 43% of all gill-breathing snails, 52% of freshwater aquatic or semiaquatic turtles, 60% of mussels, and 38% of freshwater fishes of North America (Lydeard and Mayden, 1995). Over half of the species in some of the groups mentioned are either threatened or endangered under the U.S. Endangered Species Act of 1973.

The fresh waters of the world are collectively experiencing markedly accelerating rates of degradation. Major sources of disturbance impact the biodiversity of freshwaters in many ways. Direct chemical toxicants released into surface waters and ground waters are common. Many forms of heavy metals, inorganic reducing agents, and organic compounds enter the environment and eventually fresh waters. Although some are inactivated, such as by chemical precipitation, or oxidized, many have long residue resident times. Despite dispersion and dilution in the aquatic environment, bioconcentration of both metals and organic compounds is common, by which toxicity can be increased exponentially. The biological effects of many of these compounds are unknown. Elemental and compound toxicity is increased markedly when the pollutant substances are radioactive or acidic. Increased acidity in poorly buffered fresh waters can increase concentrations and potential biotic availability and toxicity of metals, such as aluminum, that are not normally abundant in soluble, reactive states.

Nutrients, particularly phosphorus and nitrogen, can lead to well-understood enhancements of plant and other organic productivity in fresh waters. This eutrophication process often results in enhanced rates of decomposition and in chemical conditions that greatly reduce or eliminate suitable habitat for many species of plants and animals. Similar excessive loading of organic matter and enhanced rates of degradation result from organic sewage from human populations, industry, and agriculture. A further common pollutant that markedly reduces habitat availability is the suspension of finely divided organic and inert inorganic matter. Erosion and transport of such suspensions are increasing as large areas of forest and former wetland interface zones between land and water are eliminated, primarily for agricultural expansion. As a result, flow patterns of surface waters are often altered and benthic habitats of surface waters can be obliterated by sedimentation.

Changes in biodiversity in freshwater ecosystems can arise from many other disturbances. Introduction of certain competitively superior species can result in marked losses of biodiversity. The infamous example of introduction of the Nile perch into Lake Victoria of East Africa resulted in the extinction of over 200 species of its endemic cichlid fish taxa in two decades. There are many other examples of introductions of exotic plant and animal species that resulted in either direct destruction of prey or inferiorly competitive species or indirect alteration of habitats required by many species. Dense, floating macrophyte communities and other eutrophication-associated excessive plant productivity often result in deoxygenation and reduction in habitat and elimination of many plant and animal species.

Largely based on terrestrial studies, the Eltonian diversity–stability hypothesis suggested that, because of the many different traits of multiple species, ecosystems with more diverse habitats would likely have species that will survive and expand during and following an environmental disturbance and compensate for those species that are reduced by the disturbance. Therefore, a more species diverse freshwater ecosystem should be more resilient to disturbances than a less biodiverse system.

Because genetically based physiological differences and tolerances among species can be small, the individual interactive strengths of some species in a freshwater ecosystem can become saturating at high biodiversity. A point can be reached where increasing species may be functionally redundant and have reduced individual impact on the ecosystem processes. On the basis of both theoretical and experimental grounds, only a small fraction of species manipulations have strong influences on food web structure. Species redundancy implies that an appreciable functional resiliency exists in which the ecosystem can compensate in its collective metabolism and biogeochemical cycling when disturbed. Although the population dynamics become progressively less stable as the biodiversity and the number of competing species increases, biodiversity can enhance the resiliency of many community and ecosystem processes in the rate that the system metabolism returns to equilibrium states following a disturbance.

There is very little storage capacity for organic carbon within the higher trophic levels. Low residence times among the higher trophic levels results in rapid cycling of carbon and nutrients of food web components. Such rapid cycling and recycling result in a reduction in the resiliency of the higher trophic levels. Most of the storage of organic carbon occurs in the dissolved organic carbon compartment in the open water and in the particulate organic carbon deposited in the sediments. In both of these compartments, the soluble organic carbon of the pelagic areas of lakes or running water of streams and the organic carbon of the sediments, the cycling of carbon is slowed. That rate of cycling is slowed in the pelagic by the recalcitrant chemical composition of the dissolved organic carbon emanating largely from higher plants. In the sediments, cycling is further impeded by the anoxic conditions that prevail almost universally among aquatic sediments. The reduced rates of cycling and recycling result in an inherent increase in resilience stability of the ecosystem.

Any factor that influences the rates of nutrient and carbon cycling in freshwater ecosystems will influence the resilience of the ecosystem and its biodiversity to disturbances. Changing sources of organic matter, as discussed in the following, and hence bacterial metabolism and nutrient cycling thus change resilience and biotic stability.

A wealth of limnological data from a spectrum of hundreds of lake ecosystems of differing productivity suggests that with a shift in nutrient loadings, concomitant shifts occur in the development of photosynthetic producers and loadings of organic matter. During the common sequential development of lake ecosystems over long periods of time (centuries, millennia), shifts in the ratios of higher vegetation versus algal dominance can occur. Increased relative organic loading from higher vegetation results in proportionally greater loading of recalcitrant dissolved organic carbon, which can suppress nutrient cycling and increase the resilience of the ecosystem.

In addition, the development of higher vegetation in littoral and wetland combinations increases the habitat heterogeneity enormously, often by a factor of 10 or more, in comparison to lakes with limited littoral development. Species diversity nearly always increases under these circumstances by at least an order of magnitude among nearly all major groups of organisms, particularly among the lower phyla.

Greater biodiversity may have a greater collective effect by improving the capacity of ecosystems to recover from large disturbances. Reduced biodiversity increases vulnerability by reducing the total collective physiological tolerances of the community to large habitat changes. Recovery after a major or catastrophic disturbance would be slower with a reduced aggregation of residual physiological ranges within the remaining species. Recovery then must depend to a greater extent on slower fortuitous methods, such as importation of species, rather than generation from residual surviving species, or slow recolonization processes such as from remnants in resting stages or seed banks. In some cases, such as in many ponds, streams, and reservoirs in clay-rich regions where high turbidity often occurs with successive rain events, photosynthetic productivity within the water is intermittently but repeatedly suppressed. Biodiversity is likely also suppressed under these conditions or restricted to species with high reproductive potential that utilize improved conditions in periods between turbidity events.

As indicated earlier, disturbance to freshwater ecosystems can occur in many forms and to different extents. Certain perturbations can be catastrophic, such as overwhelming a lake or stream with an organic or inorganic poison in which most of the biota are eliminated. Many disturbances, however, are more gradual over long periods of time (months, year), such as nutrient enrichment, or irregularly episodic and often of short duration (days, weeks), such as severe flooding and the scouring of a section of river. Biodiversity is coupled to ecosystem stability and the type and extent of disturbances.

A model of the responses of organic productivity of lake ecosystems to changes in nutrient loading from the drainage basins has been abundantly verified by nutrient and comparative primary productivity data of phytoplankton, attached algae, and macrophytes from hundreds of lakes in different stages of ontogeny (Fig. 4). The differences in plant productivity result in very different amounts and types of chemical composition of organic matter loaded to lakes. Because of the large amounts and relative chemical recalcitrance of dissolved and particulate organic detrital sources from higher plants, decomposition of this pool is slowed. The resulting large but slow metabolism of organic detritus provides an inherent ecosystem stability that energetically buffers the organic matter reserves from rapid exploitation and cycling of nutrients contained in the organic matter. Forested watersheds, undisturbed riparian floodplains, and wetland–littoral zones are therefore important to the metabolic stability within the lakes and streams as a result of both the chemical quality and the large quantity of their inputs of organic matter. The organic carbon couplings between the drainage basin and land–water interface zones are metabolically as important to lake and stream energetic and carbon utilization stability as are the traditionally studied nutrient (phosphorus, nitrogen, silica) loading relationships to food web productivity. As the sources of organic matter change, the chemical recalcitrance to degradation can change. Alterations of rates of utilization, nutrient recycling, and energetic resiliency can ensue as a result. These changes directly affect biodiversity by altering the chemical and physical habitat characteristics.

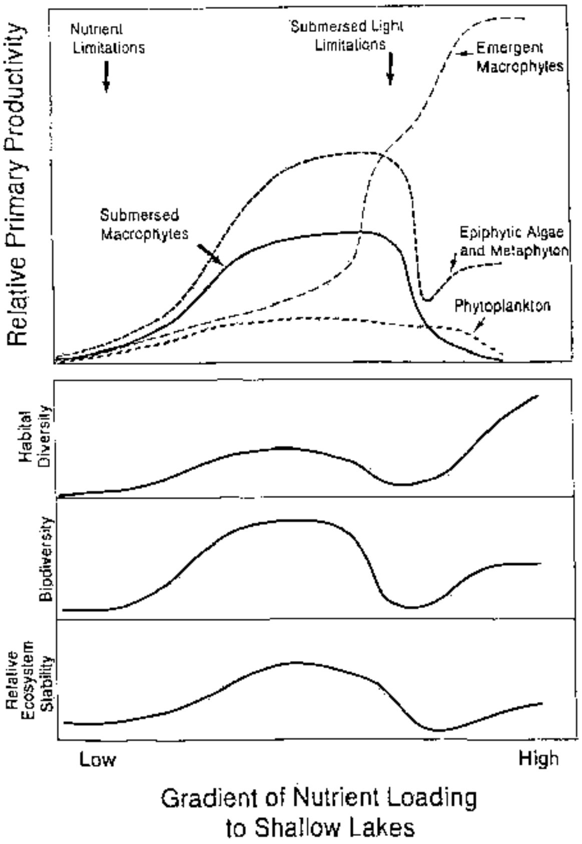


FIGURE 4 Relative changes in the primary productivity of phytoplankton, macrophytes, and attached microflora, and habitat diversity, species diversity, and ecosystem stability along a gradient of nutrient loading to lake ecosystems. (From Wetzel, 1999.)

A common disturbance to lake and reservoir ecosystems involves progressive increases in loading of nutrients. These loadings, in excess of losses to sites of temporary or permanent inactivation such as to the sediments, result in enhanced nutrient availability for phytoplankton and other autotrophs and increased rates of growth and productivity. If nutrient loading increases to an oligotrophic lake, increased productivity is rapid. Similarly, if the disturbance is brief (i.e., the duration of increased nutrient loading is relatively short), nutrient cycling is rapid, the ecosystem will recover rapidly, and productivity would be reduced proportionally to the load reduction. The return time of such an oligotrophic ecosystem is high, but the resiliency is low. Species diversity of the plankton tends to be high in oligotrophic lakes, but because of limited physical habitats, particularly associated with littoral areas and surfaces, composite biodiversity of the lake ecosystem would be low (see Fig. 4).

As nutrient loading increases, particularly among shallow lakes that predominate globally, a marked shift in the productivity occurs from the pelagic to attached surfaces associated with living aquatic plants and the particulate detritus of senescing macrophyte biomass (see Fig. 4). Under these conditions, total primary productivity and biomass increase greatly. Habitat diversity among the massively dissected surfaces of submersed aquatic plants increases exponentially, and biodiversity among attached biota would increase by at least an order of magnitude. It is likely that among all autotrophic and heterotrophic microbial groups, as well as for most of the smaller metazoans, over 90% of the total freshwater ecosystem species are sessile in association with surfaces.

As nutrient loading increases further, phytoplanktonic productivity per unit volume increases, but self-shading by the suspended algae restricts the depth of light penetration and hence the depth of the photogenic zone. Phytoplanktonic productivity per unit area declines precipitously under these eutrophic conditions of light limitation and is usually accompanied by a marked decrease in planktonic biodiversity as well (see Fig. 4). Submersed light limitations also reduce or eliminate submersed macrophyte growth and attendant surfaces for microbial growth. The elimination of photosynthesis of attached microbiota decreases collective productivity markedly. Biodiversity of the attached biota would decline precipitously as a result. The losses of attached microbial communities and their interdependent metabolism cause a massive reduction in the capacities of the lake ecosystems to retain loaded nutrients and dissolved organic compounds. Because of the great accumulation and recycling of nutrients in the sediments from a predominantly planktonic eutrophic ecosystem, the time for recovery of the water quality of a lake from a reduction of nutrient loading would be long and not proportional to load reductions. Under these conditions after the nutrient loading perturbations end, the internal nutrient loading and recycling could be greater than external nutrient loading.

Continual high nutrient loading and resulting hypereutrophic phytoplanktonic conditions generate large areas of anaerobic reducing conditions and slower rates of decomposition of organic matter. Production exceeding decomposition leads to rapid sedimentation and generation of increased shallow habitat conducive to colonization by emergent macrophytes. The high productivity of emergent macrophytes increases the proportion of lignocellulose supporting tissues that are relatively recalcitrant to rapid decomposition, particularly under reducing conditions. The collective result is markedly increased productivity and habitat diversity. However, although habitat diversity is high among very shallow wetland-dominated waters, environmental fluctuations are also more extreme than in continually submersed habitats. Quite different community structures occur and biodiversity is likely tempered as a result, though comparative data to support this statement are few.

Many descriptive and experimental studies suggest that greater biodiversity results in a higher efficiency of utilization of nutrients. Greater biodiversity likely entails more physiological diversity that can better cope with natural vagaries in environmental parameters. Competition for resources, nutrients in this specific case, is thus intensified. Efficiency of nutrient retention in an ecosystem would be maximized under conditions of greatest microbial community diversity. That microbial diversity (algae, fungi, protists, and bacteria) is maximized in the attached communities where habitat diversity and microenvironment differences are the greatest. Physical constraints of boundary layers and mucopolysaccharide matrices in which the attached microbiota live mean that movements of ions and gases to organisms living within the attached communities occur predominantly by diffusion. Nutrients, once acquired, are intensively recycled among the attached biota and conserved. Resources from external sources can then be utilized largely for new growth and reproduction. Nutrient retention is very high within the microcommunities and collectively within the attached habitats.

In certain productive lakes, piscivore consumption of planktivorous fishes can lead to enhanced development of cladoceran zooplankton. The high grazing rates of cladoceran zooplankton can result in selective reduction of larger algae for brief to moderate intervals of time. The effects of these changes among the larger forms are, however, poorly translated to smaller microbiota, and only minor changes are seen at the microalgal and bacterial levels. For example, removal of portions of the larger algae can decrease competition for nutrient and light resources. Often the smaller forms with shorter generation times increase in productivity. Nutrient recycling likely also increases, particularly as decomposition of these algae by bacteria is accelerated by protistan microconsumers (heteroflagellates, ciliates, and related organisms) and tightly retained among the microbiota. Little is known concerning changes in biodiversity under these conditions, but it appears that biodiversity among the microbiota would increase and likely compensate for the losses of larger forms by the crustacean grazing. A shift in energy fluxes occurs, here from larger planktonic forms to smaller forms with much higher turnover rates and increased rates of resource turnover. Compensatory mechanisms likely also appear in the composite biodiversity of the smaller biota. Resiliency to disturbances within the pelagic food web likely declines as a result.

See Also the Following Articles

FRESHWATER ECOSYSTEMS, HUMAN IMPACT ON • INVERTEBRATES, FRESHWATER, OVERVIEW • LAKE AND POND ECOSYSTEMS • PELAGIC ECOSYSTEMS • WETLANDS ECOSYSTEMS

Bibliography

DeAngelis, D. L. (1992). *Dynamics of Nutrient Cycling and Food Webs.* Chapman & Hall, London.

Francko, D. A., and Wetzel, R. G. (1983). *To Quench Our Thirst: Present and Future Freshwater Resources of the United States.* University of Michigan Press, Ann Arbor.

Lydeard, C., and Mayden, R. L. (1995). A diverse and endangered aquatic ecosystem of the southeast United States. *Conservation Biol.* 9, 800–805.

Tilman, D. (1996). Biodiversity: Population versus ecosystem stability. *Ecology* 77, 350–363.

Wetzel, R. G. (1983). *Limnology*, 2nd ed. Saunders College Publ., Philadelphia.

Wetzel, R. G. (1990). Land–water interfaces: Metabolic and limnological regulators. *Verh. Int. Verein. Limnol.* 24, 6–24.

Wetzel, R. G. (1995). Death, detritus, and energy flow in aquatic ecosystems. *Freshwater Biol.* 33, 83–89.

Wetzel, R. G. (1999). Biodiversity and shifting energetic stability within freshwater ecosystems. *Arch. Hydrobiol. Beih. Ergebn. Limnol.*, 54, 19–32.

Wetzel, R. G. (2000). *Limnology: Lake and River Ecosystems.* Academic Press, San Diego

FRESHWATER ECOSYSTEMS, HUMAN IMPACT ON

Kaj Sand-Jensen
University of Copenhagen

GLOSSARY

denitrification Bacterial nitrate respiration of organic matter to elemental nitrogen under oxygen-free conditions in soils and aquatic sediments.

detritus Dead organic matter.

oxic and anoxic conditions Environmental conditions with and without oxygen, respectively.

Red Lists Compilations of recently extinct, threatened, and vulnerable species for a country or a larger region.

Secchi-depth Measurement of the transparency of water by determining the depth at which a white Secchi-disc (20-cm diameter) lowered from the surface disappears out of the sight of the observer.

FRESHWATERS ARE INHABITED BY A REMARKABLE VARIETY OF SPECIES from most taxonomic groups evolved on Earth, but they are also a scene of human opportunities, requirements, and conflicts. Strong human impacts have restricted the number and variety of freshwaters and polluted and disturbed many others. This article discusses the range of human impacts and their consequences for species richness and functional diversity of individual lakes and streams. Special emphasis is placed on freshwaters in Europe and North America and their intensely cultivated regions. Biological invasions are discussed briefly; climatic changes are not dealt with because of a lack of general predictions.

I. FRESHWATERS: OPPORTUNITIES AND CONFLICTS

Why are we so interested in freshwater and natural freshwater ecosystems? The answer is obvious—we cannot live without it! Freshwater is essential to our personal daily life and the production of vital food and industrial products.

We use freshwater to drink and to prepare and cook our food. We use freshwater to wash ourselves and our clothes and to remove waste products. Securing clean tap water and good sanitary conditions through installation of water closets and sewer pipelines has been important in reducing infectious diseases and infant mortality. Daily life also includes the enjoyment of natural lakes and streams, which provide variation in the landscape and opportunities for bathing, fishing, sailing, and observations of animal and plant life. Therefore, it is of high priority to all human beings that clean water

Encyclopedia of Biodiversity, Volume 3

and unspoiled freshwater habitats are available in great amounts.

Freshwater is also required for irrigation of agricultural crops, and it is used in many industrial processes and to transport waste products. Because of the magnitude of these uses, the pressure on groundwater and surface waters is immense in regions with dense populations and intense agriculture and industry. Unfortunately, many of the uses deteriorate water quality, thereby affecting water use for other purposes. The growing number of uses and users generates conflicts among neighboring countries and different user interests, and further exacerbates the dilemma of rising demand and the need and desire for high-quality aquatic habitats. Overall, freshwater ecosystems have been more restricted, manipulated, and polluted than any other ecosystem on Earth, and conflicts will intensify in the future in the light of growing human populations and contemporary environmental trends.

II. THE SMALL, BUT NUMEROUS FRESHWATER HABITATS

Earth has much water. A deep, continuous ocean covers 70% of Earth's surface and contains 97% of all water. Only 0.65% of all water is found as freshwater on the continents, and most of this is groundwater (0.62%). Even smaller proportions are found in streams (0.0001%) and lakes (0.017%).

Lakes and streams cover a variable proportion of the land surface. In dry regions, surface waters cover less than 0.1% of the land surface. In wet environments such as the tundra, the boreal forest, and the rain forest, plenty of water is found in shallow pools, lakes, and streams, which may occupy 5–10% of the land surface. A map of Great Britain reveals that surface waters cover less than 0.1% of the land in the south and more than 5% in the rainy uplands of Wales and northwest Scotland; the average water coverage is 1–2%.

Because inland waters are mostly small, shallow, and numerous, they form an intimate contact to the terrestrial environment. This contact has been essential for the historic and contemporary development of species and life stages of amphibious plants and animals. Despite the small volume of lakes and streams compared to the ocean's volume, freshwater environments have been a vital platform for the evolution of many lines of algae, plants, and animals.

It is possible to quantify the contact between terrestrial and freshwater environments. In a small country like Denmark, inland waters occupy 1.7% of the terrestrial area, and on average 1 km^2 of land area is in contact with 3.6 km shoreline of freshwater lakes and streams and 0.2 km of the sea, according to estimates from 1:25,000 maps. Before agriculture removed numerous shallow lakes, pools, and streams, the freshwater contact was probably two or three times longer.

The greater freshwater than marine contact to land is more striking for the world, because land areas are joined in large continents. As land-masses grow in size, the contact zone to the sea increases only with the periphery and, thus, with the square root of the land area, whereas the contact zone with inland waters approximately increases in proportion to the land area. For the continents, the estimated contact zone is 100–1000 times longer between land and freshwater than between land and sea.

The transition between land and freshwater is gradual, gentle, and suitable for organisms, because physical forces and disturbance are weaker than in the transition from land to sea. The transition zone from small lakes and streams to land has a closed vegetation cover; that of the oceanic coasts faces strong winds, waves, and moving sand with only scattered vegetation. It has been much easier for organisms to cross between land and freshwaters than between land and the sea. The many new freshwater bodies that are formed during glaciations or appear transiently during wet periods also offer opportunities for the emigration and development of new species without the intense competition and predation caused by well-established species, as is typical of the sea. On the other hand, shallow lakes disappear after only a few thousand years, because they are filled with particles eroded from the land and organic matter produced in the lake. In essence, freshwaters provide many opportunities, which constantly come and go, but they lack the long stability of the sea that has lasted for billions of years.

This scenario should have stimulated the selection of species with rapid evolution and efficient means of dispersal allowing them to colonize new freshwaters as their original habitats disappear. Indeed, apart from species of fish, molluscs, and crustaceans associated mainly with large ancient lakes (e.g., Lakes Baikal, Titicaca, Victoria, and Tanganyika) and large ancient river systems (e.g., the Danube, rivers of southeastern North America, the Amazonas, and other tropical rivers), most freshwater species are widespread within and even among continents. Freshwater microorganisms among the bacteria, algae, and protozoans are generally both locally abundant and cosmopolitan. The same species of cyanobacteria and microalgae live in extreme and rare freshwater habitats such as hot springs or highly acidic pools in distant locations on different continents and oceanic islands. Most aquatic rooted plants are also widespread, though a few species are endemic because underwater dispersal restricts their spread.

III. SPECIES EVOLUTION AND RICHNESS IN FRESHWATER

As a result of the intimate contact, the long coastline, and the suitable transition zone, there has been a lively exchange of species from freshwaters to terrestrial environments and back again. Plants have evolved from a special group of freshwater green algae (Charophyceae, Coleochaetales) and diversified under the highly variable conditions on land. Insects have evolved from groups of arthropods in the transition between freshwater and land. Amphibians and reptiles have evolved from special fish in freshwater and brackish wetlands. Many evolutionary lines have been followed at different times and places.

Freshwaters have formed a corridor for the two-way dispersal of organisms between land and sea. Freshwaters also share many taxonomic groups of algae, plants, and animals with the terrestrial environment and the sea, while divergences are stronger between the land and the sea. Freshwater environments are surrounded by large surface areas of terrestrial and oceanic environments, which promotes the emigration of marine and terrestrial species and their adaptation to freshwater environments.

Among major groups of land plants, several have secondarily returned to freshwaters, though few have reached the sea. Among liverworts, true mosses, bog mosses, horsetails, and ferns there are many freshwater species, but no marine representatives. Among flowering plants, 1000–1400 truly aquatic freshwater species have been described from many plant families, illustrating that the secondary return from land to freshwater has occurred independently and repeatedly. The return process continues today, with more than 4000 species living an amphibious double-life in the transition between land and freshwater. The sea includes only about 60 species of flowering plants with a restricted taxonomic diversity.

The diverse freshwater insects (>45,000 species) live a double-life as eggs, larvae, and pupae in water and as flying imagines dispersing and mating on land. Mayflies live for only a few days on land without taking food, while adult dragonflies live a long and active terrestrial life. Freshwater insects, therefore, require a suitable environmental quality in both freshwater and adjacent terrestrial environments for their sustained survival and development, though most evaluations of habitat quality in relation to species composition and richness of insects focus only on the aquatic zone.

No phylum—the highest taxonomic entity of organism below the kingdom—is found exclusively in freshwaters, but all large phyla apart from the echinoderms are represented. Virtually all types of photosynthetic organisms, including cyanobacteria, the great variety of algae, mosses, and vascular plants, are represented in freshwaters (Table I). In contrast, the sea has very few species of gold algae, yellow-green algae, plants, and insects, and it lacks some groups of green algae (e.g., desmids and stoneworts, known as characeans). The sea outnumbers freshwaters only with respect to photosynthetic species among red algae, brown algae, and dinoflagellates.

The three main determinants of species richness for a specific type of environment are surface area, habitat heterogeneity, and the time history in terms of durability and evolutionary development. Considering that freshwaters cover only a few percent of the surface area relative to that of the sea and the terrestrial environment, and considering that inland surface waters mostly have a short life-time, the freshwater biodiversity is surprisingly high and must arise from the high habitat

TABLE I

Estimated Number of Algal Species in Freshwaters and Total Number of All Species in Soils, Freshwaters, Brackish Waters, and the Sea[a]

Taxon	Freshwater species	All species
Cyanophyta (blue-green algae)	1500	2000
Rhodophyta (red algae)	150	5000**
Chrysophyceae (gold algae)	1900*	2000
Xanthophyceae (yellow-green algae)	550*	600
Bacillariophyceae (diatoms)	5000	10,000
Phaeophyceae (brown algae)	Few	2000**
Cryptophyta	100	200
Dinophyta (dinoflagellates)	200	2000**
Euglenophyta	700*	800
Chlorophyta (green algae)	7000*	8000
Zygnematophyceae (e.g., desmids)	6000*	6000
Charophyceae (stoneworts)	80	80

[a] Values are presented for major taxonomic groups. Mainly freshwater groups (*) and marine groups (**) are marked.

variability among individually confined water bodies and the suitable conditions for emigration and establishment of species from the large, adjacent terrestrial and marine environments.

Taking the vertebrates as an example, there are about 10,000 named species of freshwater fish and 15,000 species of marine fish. Among birds and mammals, virtually all depend on freshwater bodies for drinking, but a large proportion of the species also are dependent on freshwater lakes, streams, and wetlands for breeding and feeding. In Europe, about 25% of the bird species and 11% of the mammal species live their entire life or part of it in freshwaters.

A. Species in Freshwater Sediments

A global overview of organisms associated with freshwater sediments yields ca. 175,000 described species (Table II), but the true number of species is much higher than this. The most speciose groups in freshwater sediments are the invertebrates and especially the insects, nematodes, and crustaceans. Among nematodes and rotifers living between the particles in freshwater sediments, there are probably many thousands of undescribed species.

Compilation of local species richness and taxonomic diversity (see Table II) yields equally high, or often even higher, values in freshwater environments than in marine or terrestrial environments. There are no systematic differences in local species richness between lakes and streams, but upstream sections of streams lack the variety of phytoplankton, zooplankton, and fish that are typical of open waters, and instead support a greater species richness in the surface sediments and in the deep sediments below the streambed. The downstream parts of streams have a greater resemblance to lakes, so the similarity of species composition and richness will depend on whether entire stream systems or just certain stream sections are compared with the lakes. The biota in porous groundwaters is much deprived in species due to lack of light, degradable organic matter, and dissolved oxygen. Specialized species of protozoans and small invertebrates live here in local species numbers that are typically 10 to 20-fold lower than those encountered in lakes and streams (see Table II).

TABLE II

Number of Species Described Globally from Freshwater Sediments and the Typical Local Range of Species in Lakes, Streams, and Groundwaters[a]

Taxon	Global	Lake/stream	Groundwater
Bacteria	>10,000	>1000	>100
Fungi	600	50–300	0–10
Algae	14,000	0–1000	0
Plants	1000	0–100	0
Protozoans	<10,000	100–800	0–20
Crustaceans	8000	25–150	5–60
Insects	45,000	50–300	0–10
Molluscs	4000	0–50	0–10
Other invertebrates	>12,000	30–70	5–70

[a] Data compiled by Palmer *et al.* (1997).

IV. HUMAN IMPACTS ON FRESHWATER ECOSYSTEMS

When evaluating the human impacts on freshwaters, we tend to concentrate on the numerous examples of water pollution. Over the last 150 years, the environmental issues have gradually changed as new problems have appeared and become recognized. Since the mid-1800s, organic pollution of streams and lakes with organic wastes from households and domestic animals has been of major concern in Europe and North America. Since 1945, the focus has been on cultural eutrophication of inland and coastal waters with nitrogen and phosphorus from agriculture, towns, and industries. After 1970, acidification of inland waters came onto the agenda due to increasing concentrations of sulfuric and nitric acids in the precipitation and changes in land use. The latest chemical concerns include trace metals and an enormous range of synthetic organic compounds of largely unknown behavior and ecological effect.

Many physical changes in the catchments are, however, much more important to the existence, environmental quality, and biodiversity of surface waters than the direct water pollution. The most significant influence on terrestrial and freshwater environments is the removal of natural vegetation and the cultivation of land, which lead to immediate, profound alterations of the hydrology and nutrient cycling. These alterations are grossly enhanced when soils are drained and streams are canalized. The intimate linkage between natural wetlands and streams, which has been important for the evolution and contemporary diversity of plant and animal life, is also disrupted when surplus water directly flows to the stream through drain pipes rather than slowly percolating through the wet, sponge-like organic soils.

The impact of biological invasions is viewed as being much more critical in North America than in Europe, probably because North America has been disturbed more recently by the agricultural settlement of European colonists, and because it includes a greater variety of endemic fish and large invertebrates that survived the latest glaciations in refugia in the U.S. Southeast and Mexico. Allan and Flecker (1993) discussed the serious threats of biological invasions to biodiversity in temperate (e.g., New Zealand) and tropical countries.

Most tropical countries now face the unregulated impact from all pollution sources. The most widespread problem appears to be the rising organic pollution of streams and lakes by untreated domestic sewage. However, in some regions heavy pollution takes place from (1) extensive application of pesticides and nitrogen in plantations of bananas, cocoa, and oil palms, (2) large oil spills (e.g., Ecuador, Venezuela, Nigeria), (3) acidification and pollution with heavy metals (e.g., copper, mercury) from mining areas and tanneries, and (4) outlet of phenolics from the timber industry. A fundamental problem is the erosion of soils deriving from the massive deforestation of tropical rain forests. Many streams have become chocolate colored from the heavy load of solids eroded by the strong rainfall. Erosion is a severe problem in agriculture, which loses the fine, fertile top-soils. It also causes environmental problems as hydroelectric dams are quickly filled with sediments, and when suspended loads in streams prevent light penetration and thereby the growth of algae, which are food for many invertebrates and fish. Furthermore, stream bottoms get clogged with fine sediments that destroy the habitats of invertebrates and the spawning banks of fish.

In many tropical countries, physical alterations of streams are still relatively few. In Ecuador, for example, streams still run in natural streambeds with meanders and rapids surrounded by strips of riparian vegetation even in densely populated areas. The hydrological cycle has been altered, however, since water is removed for irrigation, leading to artificially low discharge and current velocity during the dry season. As a consequence, the stream biota is impoverished due to oxygen depletion and smothering of the streambed following organic pollution.

A. Restriction of Area and Variability of Freshwater Habitats

The scene of physical changes and areal reductions of freshwater ecosystems changes among countries and with time. Changes of freshwater ecosystems have been greatest over the last 100–200 years in densely populated countries and in regions with intense agriculture and industry. Thus, the combination of population density, resource use, and powers of technology is a suitable measure of human impact on the biosphere in general and the freshwaters in particular.

Petts (1984) defined four phases in the recent era of river modification in Europe. Phase 1, from 1750 to 1900, includes ambitious regulations of the major rivers for the purpose of navigation, flood control, and cultivation of the river valley. Phase 2, from 1900 to 1950, marks the first major technological period, during which large dams and power plants were built across major rivers. In many European lowlands, extensive drainage of wetlands and shallow lakes and channelization of streams took place, and continuous management of the dimensions of stream channels and cutting of aquatic plants were initiated. Management intensified in phase 3 from 1950 to 1980 by the use of specialized machines. During the recent phase 4, from 1980 onward, the intensity of regulation works and dam building has gone down, because most watercourses have already been exploited and public resistance has increased because of rising environmental concerns. However, most countries of the world are still in the most exploitive phase 3.

An overview of the 69 major rivers in Europe and Russia shows that most of them are strongly (40) or moderately (10) affected by dams and regulations, whereas only 19 rivers located in Arctic and northern boreal regions have remained relatively unaffected. River regulation has been undertaken to the greatest extent in western and southern Europe. In Belgium, Denmark, England, and Wales, the percentage of river reaches that are still in a natural state is less than 20%. In Estonia, Norway, and Poland, the rivers still have 70–100% of their reaches in a natural state.

In southern Canada and the United States, rivers are heavily regulated; those in northern Canada and Alaska have remained relatively pristine. The history of Willamette River in Oregon is an example of how expansion of agriculture and construction of 11 dams have transformed a complex multi-channeled river into a simple one- or two-channeled river. Over a 25-km-long stretch in the floodplain, the length of the shoreline has declined from 250 km in 1854 to only 64 km in 1967.

In addition to the shortening of rivers and streams, watercourses have become more uniform and physically disturbed. When streams are channelized, the natural variability in depth, width, current velocity, and sediment composition disappears between straight reaches and meanders and between riffles and pools. The natural dynamics of the flow channel, characterized by spatial variations of erosion and sedimentation, formation of new meanders, and the cut-off of oxbow lakes, also vanishes. Drainage of the floodplain reduces its storage capacity of water and the ability to buffer storm surge in the rivers. Moreover, drainage reduces baseflow during dry periods. Peak discharge during rainy periods introduces high physical stress and erosion in the stream channel, and low baseflow during drought increases the risk of high water temperatures, oxygen deficiency, and insufficient currents for the well-being of invertebrates and fish. All three major changes—(1) smaller area and length of the streams, (2) lower habitat diversity, and (3) greatly fluctuating flow and higher physical, oxygen, and temperature stresses—will drastically reduce species number and population density of plants and animals.

Numerous shallow lakes and ponds in European farmland have disappeared from drainage and filling-in. Many ponds were originally dug to extract calcareous sediments for the liming of acidic fields, clay for bricks, and sand and gravel for construction works and to provide watering places for cattle. In some regions more than half of the ponds that were present 150 years ago are now gone. As a result, several species of amphibians and large insects that preferred the warm, shallow pond waters free of predatory fish have declined.

Large, deep lakes have experienced less physical disturbance and restriction of areal cover than streams and ponds. Exceptions are artificial reservoirs, which face great fluctuations in water level over the seasons. In Europe there are now more than 10,000 reservoirs, and about 4000 large reservoirs have dams higher than 15 m. Dams have broken the natural connection and dynamics of watercourses for the movement of water, solutes, and living organisms. On the other hand, reservoirs have generated habitats for lake organisms and new recreational opportunities.

B. Changing Direction: Restoring and Creating New Freshwater Habitats

Although areal reduction and the manipulation of inland waters continue in most countries, public and political priorities have changed in some wealthy countries with the goal of ameliorating some of the environmental damages caused by intensive agricultural and industrial practices.

Norway experienced serious demonstrations when a hydroelectric power plant was built across the former pristine Alta River. In Spain, the dam across the Esla has remained unused, because the inhabitants of the valley to be expropriated have refused to move. In Iceland, local "farmers" stopped early attempts to build a dam across the famous Laxa River. In Denmark, new laws now give higher priority to the development of diverse landscapes, diverse biological communities, and trout fishing. After water quality had been improved following the extensive purification of domestic sewage, new management procedures and structural changes have been implemented in many stream reaches to improve habitat variability, to prevent catastrophic disturbance, and to strengthen the linkages within the stream course and between the stream and the floodplain. These Danish initiatives include: (1) removing all weirs and dams blocking the natural migration of animals; (2) bringing the streams that had been directed underground in culverts to the open surface; (3) reestablishing the variation in width, depth, and substratum between riffles and pools; (4) re-creating the former meandering patterns of straightened channels; and (5) establishing strips of natural vegetation free of pesticide spraying along the streams.

Another initiative in northern Europe has been to restore shallow lakes and ponds that were facing disappearance due to lack of water, filling in, or overgrowth by littoral vegetation. New ponds have been created in farmlands with the purpose of regaining the patchwork of small, suitable biotopes for aquatic plants, insects, and amphibians. Populations of several threatened species of frogs and newts have since increased thanks to reestablishment of suitable ponds.

C. Changes of the Hydrological Cycle

It is often over-looked how much the hydrological cycle has been changed by humans and the environmental implications of these alterations. Archeology, paleolimnology, and historic studies of old maps and written records in Europe reveal that the groundwater level was much higher and there used to be many more open waters in the past compared to today. Previously, water moved slowly and diffusively through the large water reservoirs in the soil and the wetlands before reaching streams and lakes. Following cultivation of the soil and establishment of ditches and drains, the water has taken a more direct and rapid course from the fields to the streams, resulting in huge variations in discharge. Water for irrigation of crops and domestic supply to metropolitan cities is abstracted from groundwaters and surface waters, and this water demand has markedly increased the risk of wetlands drying up and critically low flows developing in the streams during dry periods.

Breaking the long contact between water and soil has dramatically reduced the natural purification of water by preventing the removal of surplus nitrogen and phosphorus through binding to soil particles, incorporation in new organic material, precipitation as phosphate minerals, and denitrification. In addition to the increased cultivation of crops, application of fertilizers, and discharge of sewage, changes in the hydrological cycle have resulted in the explosive eutrophication of inland and coastal waters. A few countries have tried to reestablish some of the wetlands that have been lost so extensively, thereby reinstalling the cleaning sponge between cultivated soils and streams. In most other countries, however, wetlands continue to disappear at a high rate. Spain, for example, has lost two-thirds of its wetlands since 1965 with the support of European Union subsidies to farmers.

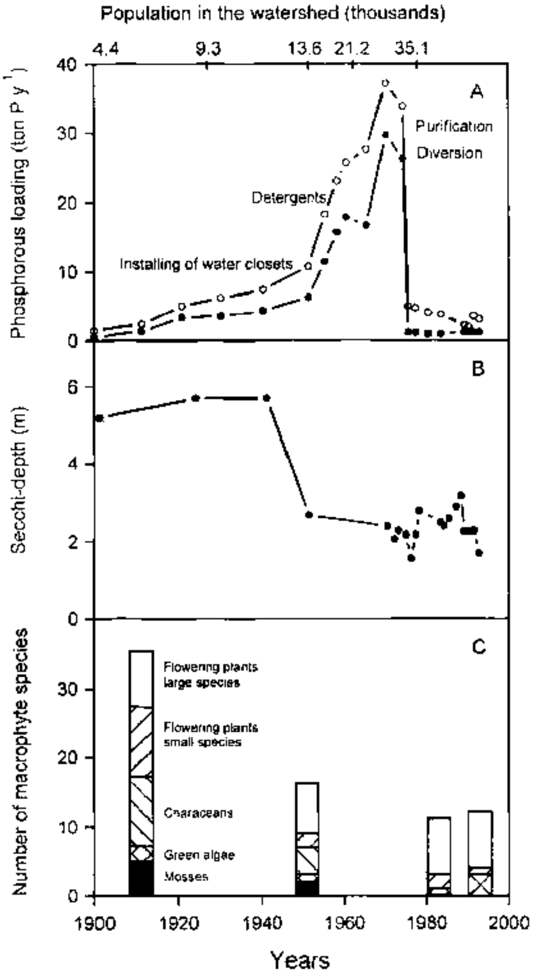


FIGURE 1 Historical development of eutrophication in Lake Fure in Denmark from 1900 to 1993. (A) Annual phosphorus loading, showing the early installation of sewers, the application of detergents, and the recent diversion and tertiary treatment of domestic sewage. Open symbols mark total loading, whereas closed symbols mark sewage loading. (B) Decline of Secchi-depth over time. (C) Number of submerged macrophytes among mosses, characeans, and small and large rooted plants on four occasions. (After Sand-Jensen and Pedersen, 1997.)

D. Organic Pollution of Inland Waters

The discharge of organic wastes from towns and from livestock operations creates the classic pollution problem of inland waters. Human wastes are derived from bathing, cooking, laundry, and the flushing of feces and urine in lavatories, and discharges from agriculture include manure from the animals and food spills. Intense organic pollution can also occur from breweries, dairy factories, slaughterhouses, sugar refineries, and countless other industrial sources. In regions where water quality otherwise is high, the farming of fish, prawn, and crayfish can be common. The trout farms of northern Europe mainly use cold spring water of superb quality, but these farms in turn pass on substantial organic pollution from food spills and fish excreta.

With the increase in domesticated animals, human populations, and the installation of water closets and sewers, there has been a profound increase of organic loading of inland waters in Europe and North America from 1850 to 1950. In several cases, the external loading of lakes has increased more than 10-fold (Fig. 1). Substantial purification of domestic sewage has been established over the last 50 years. Meanwhile, wet slurry from burgeoning animal farms has presented a new source of pollution. In poor countries of the world, organic pollution has continued to grow and few attempts have been made to ameliorate the problem through the use of natural wetlands with self-purification capacities or the construction of costly sewage treatment plants.

The organic pollution of streams increases with the density of human beings and animal livestock and the consumption of oxygen by organic matter in the water. An acceptable water quality can be reached by extensive purification, as shown by a 10-fold reduction of the biological oxygen demand at the outlet from some sewage treatment plants and a 3-fold reduction in the stream water in Denmark from 1975 to 1995. Otherwise, the European inland waters with the best water quality in terms of low oxygen demands and high oxygen concentrations are found in the water-rich regions in the north, where population densities are low and large areas remain uncultivated. In contrast, streams with the poorest water quality are located in intensely cultivated regions of middle and southern Europe, where purification is insufficient and water is in short supply during the summer. To illustrate the differences in environmental state, oxygen conditions are regarded as good in 97–99% of the streams in Scotland and Iceland and 64–77% of the streams in Wales, but in less than 20% of the streams in Belgium, Poland, and countries in the Balkans. Most European countries have experienced an overall decline of organic pollution in the 1980s and 1990s.

E. Agricultural and Industrial Chemicals

An immense variety of chemical products are being manufactured, used, and released by agricultural and industrial activity. Some inorganic compounds include acids, alkalis, ammonia, chlorine, radio nuclides, and heavy metals (e.g., cadmium, copper, iron, mercury, and zinc). Organic compounds are grouped under different names such as chlorinated hydrocarbons, hydrocarbons, pesticides, phthalates, and phenolic compounds. Perhaps 20,000–50,000 substances are manufactured or applied within industrial countries, and a few thousand are added each year. In most European countries, between 120 and 530 active pesticides are approved for agricultural use today. The annual usage is usually between 1 and 14 kg per hectare (ha) of arable land. Substances that are applied within a country are also detected (or their degradation products) when tested for in surface waters and groundwaters. The most common pesticides in groundwaters (i.e., atrazine, desethylatrazin, lindane, and simazine) are often found in concentrations exceeding the maximum allowable threshold (0.1 $\mu g\ L^{-1}$ in the European Union).

A major obstacle to pollution control is the need to first recognize the presence of potentially harmful substances in potentially harmful concentrations in relevant ecological situations. This requires a lot of money, appropriate chemical skills, and advanced analytical methods to conduct an adequate survey of the distribution and concentrations of just a small number of these trace organic chemicals. It is then even harder to evaluate the biological consequences under natural conditions. As more traditional and obvious pollution problems in developed countries are stabilized or reduced, these organic substances may perhaps become the key pollution problem in many water bodies. The pollution effects may become more apparent as the application of these new organic compounds spreads and they accumulate over time in groundwaters. However, because their biological effects are subtle, chronic, and extremely costly to verify in the complex blend of numerous compounds, living organisms, and environments, the only effective solution to the pollution problem is not to discharge the pollutants at all, or reduce their toxicity, magnitude, and rate of application as far as possible. In less-developed countries, it will be impossible to monitor the release and biological effects of these trace organic compounds. Some frightening pollution events have been reported from the heavy misuse of pesticides in tropical crop production.

F. Cultural Eutrophication

Cultural eutrophication—predominantly due to increasing loads of nitrogen and phosphorus—leads to profound changes in the composition, biomass, and productivity of algae and plants. Lake eutrophication results in phytoplankton blooms, untransparent water, and oxygen deficiency. Eutrophication spoils the quality of bathing water and threatens the survival of bottom animals and fish. Algal blooms can include toxic algae and thereby harm animal life and become a public health risk.

Eutrophication of streams can also enhance the growth of attached macroalgae and flowering plants. Most European lowland streams have long passed the threshold at which nutrient concentrations limit plant growth. Streams in sparsely populated regions of northern Europe and Canada can still hold such low concentrations that plant growth is enhanced at sites of elevated nutrient input. Stimulation of algal growth by eutrophication is also very important in most tropical streams.

Nutrient input to watercourses has increased dramatically during the last 150 years, and it has intensified over the last decades. The sources of nitrogen and phosphorus input include (1) towns and industries, (2) scattered settlements, (3) agriculture, and (4) a background input deriving from precipitation and runoff from uncultivated areas. Input from towns and agriculture usually dominates the overall nutrient budget, but all four sources have increased because of anthropogenic impacts.

The classic examples of lake eutrophication have been documented in the vicinity of metropolitan towns. In the case of Lake Fure close to Copenhagen, annual phosphorus input has increased 30-fold from 1900 to 1969 due to an eightfold increase of the population density in the catchment, the installation of sewers, and the use of phosphorus-rich detergents (see Fig. 1).

Agriculture is the other major source of nutrient loading. Paleolimnological studies document the increase of accumulation rates of mineral particles, phosphorus, and organic matter in lake sediments due to increased erosion and runoff following the early cultivation of watersheds. Though nutrient input with domestic sewage has recently declined in some industrial countries thanks to tertiary sewage treatment, agricul-

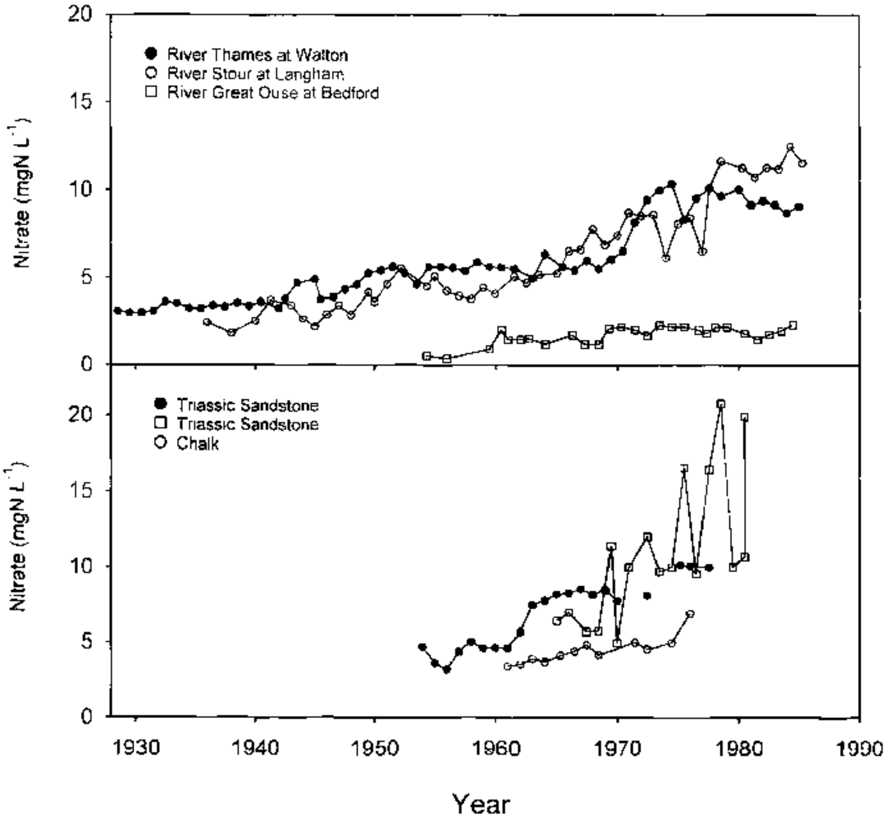


FIGURE 2 Rise of nitrate concentrations in selected English rivers and groundwater reservoirs between 1920 and 1990. (Data compiled by Heathwaite, Johnes, and Peters, 1996, and adapted from Moss, 1998.)

tural sources of nitrogen have either leveled off or continued to increase because of the heavy application of fertilizers. A large proportion (50–80%) of added nitrogen is washed out as nitrate from the soils or released as gaseous ammonia to the atmosphere. Measurements in England between 1920 and 1990 show a 2- to 6-fold increase in nitrate concentrations in a series of streams and groundwater reservoirs over periods of variable duration (Fig. 2).

An assessment of European streams reveals that nitrate concentrations have increased from 1970 to 1990 because of the strong agricultural input, while phosphate concentrations have declined thanks to better cleaning of domestic and industrial sewage (Kristensen and Hansen, 1994). With unregulated application of phosphorus fertilizers, phosphorus pools in agricultural soils have increased, and phosphorus will eventually get lost to inland waters by erosion of soil particles and by leaching as the binding capacity to soil minerals is surpassed. This development represents a new risk to the control of lake eutrophication in countries attempting to control phosphorus through tertiary treatment of domestic sewage. In streams exposed to strong human impact, median concentrations of nitrogen and phosphorus are typically 10-fold higher than in the few pristine streams in mountain ranges and forest regions. However, even oligotrophic waters in uncultivated areas have experienced a 2- to 6-fold enrichment of nitrogen through atmospheric deposition since 1950 in most regions of Europe and North America.

G. Anthropogenic Acidification

Anthropogenic acidification of precipitation and surface waters commenced with the Industrial Revolution, but accelerated in Europe and North America after the 1950s. Burning of fossil fuels releases sulfur oxides, which are converted to sulfuric acid in the atmosphere. High combustion temperatures in car engines and power plants release nitrogen oxides, which are converted to nitric acid. Although the use of cleaner fuel gases reduced sulfur emission in the 1980s and 1990s, emission of nitrogen oxides from car traffic has continued to increase.

Other conditions may contribute to the acidification of inland waters. Drainage and mining can expose metal sulfides (notably pyrite) to oxygen, starting one of the strongest acidifying processes that has sulfuric acid and ochre as the end products. The large input of ammonium with precipitation and wet slurry can acidify the soils and the groundwater through conversion of ammonium to nitrate. Cultivation of the land and changes from deciduous to conifer forests also affect acidification in the catchment by influencing the deposition of sulfur oxides and ammonia in the plant canopies and the strength of acidifying or buffering processes by chemical weathering and biological transformations.

Acidification is substantial in the northeastern United States and southeastern Canada. Acidified regions of Europe include large parts of Finland, Norway, and Sweden, mountains in Great Britain, Poland, and the Czech Republic, and carbonate-poor, well-leached sandy soils in Belgium, Denmark, and the Netherlands. In Norway and Sweden, acidification represents a significant national problem as thousands of inland waters have been acidified and lost their resident invertebrate species and fish. Vast forested areas have also been acidified to the extent that species diversity and elemental cycling have been grossly disturbed.

The biological effect of acidification is in part due to the direct reduction of pH, which is intolerable to many organisms. Associated effects appear from an altered ionic balance, high metal concentrations, and immobilization of vital nutrients. Acidification has also affected a large proportion of oligotrophic softwaters in uncultivated, remote areas. Many freshwater habitats that have escaped drainage and eutrophication have instead been acidified.

The best way to reduce acidification is to lower the emission of sulfur dioxide and nitrogen oxides through national and international agreements rather than trying to reverse the problem locally. Thus, the reduction of sulfur emissions appears to have reversed acidification in many regions in the 1990s. Also, a remarkable national scheme in Sweden to lime sensitive lakes has had significant local effects.

V. BIOLOGICAL QUALITY OF FRESHWATER ECOSYSTEMS

Anthropogenic impacts on freshwater ecosystems lead to loss of lakes and streams, changes in the hydrological cycle, physical disturbance, and water pollution. Streams have presumably experienced the entire range of physical and chemical influences, whereas major lakes mostly have experienced acidification, eutrophication, and pollution by organic matter and industrial chemicals.

The resulting changes of species composition, species richness, and elemental cycles in freshwater ecosystems are complex, have several causes, and can be evaluated at different levels. Major changes have taken place in almost every stream due to regulation, management, and pollution, and in many regions it is impossible to find pristine streams to use as a reference for baseline conditions and high biological quality.

The natural variability among lake types has dropped. Historic and paleolimnological studies of Danish lakes, for example, show that the former full range of lake types from oligotrophy to eutrophy has been restricted owing to widespread eutrophication such that there are very few lakes left that can meet the international criteria of oligotrophy. Though probably more than 80% of Danish lakes, belonging to the oligotrophic, mesotrophic, and eutrophic categories, had a rich submerged vegetation 100–150 years ago, less than 15% of all lakes today have sufficiently clear water to permit submerged plant growth. With the exception of Finland, Norway, Scotland, and Sweden, oligotrophic, clear-water lakes have become rare across in Europe.

Less variation in lake and stream types will restrict species richness and, in particular, reduce population densities of those species that prefer undisturbed, oligotrophic habitats. Likewise, in the European terrestrial vegetation, many oligotrophic and disturbance-sensitive species characteristic of heathers, *Sphagnum* bogs, and nutrient-poor grasslands have become rare and are threatened by local or regional extinction. In many individual lakes and streams we can document profound alterations of species composition and abundance over time. Recent changes are small, however, if we evaluate only whether the individual species still exist within the national borders, or have survived globally. There is a high probability that refugia of undisturbed lakes and stream reaches still exist somewhere, and that they support small populations of those species that otherwise have declined dramatically in geographical range size and local abundance. Species extinction is a slow process, and long before the last individual finally dies, the species has lost its significance in the ecosystem and in our perception of nature.

A. Decline of Freshwater Species in Europe and North America

A large proportion of extinct, threatened, and rare species in Europe and North America live in freshwaters. Species in streams and ponds have been most strongly affected by human impact. In North America, 11–14% of the mainly terrestrial birds, mammals, and reptiles are among the extinct, threatened, and rare species (Table III). Among fresh-water amphibians, fish, crayfish, and mussels, the percentages are particularly high (28–73%). No less than 103 species of North American freshwater fish are classified as endangered, 114 are threatened, and 147 deserve special attention. These species represent about one-third of all fish species. Twenty-seven species have gone extinct over the last 100 years from habitat loss, chemical pollution, introduction of exotic species, hybridization, and overfishing (Allan, 1995).

A similar assessment is not available for Europe. However, the European status of freshwater species is probably worse because of a long history of strong human impact. Europe is relatively poor in species of freshwater fish (ca. 250) compared with North America (ca. 850), Africa (1800), and the Amazon region (2000 species). Species numbers are particularly low in northern Europe (e.g., 38 species in Denmark), while they are much higher in middle and southern Europe. The Volga and the Danube include 60–70 species each, or 25% of the entire European fish fauna. In European countries about one-third of all fish species are on the IUCN Red List, much like in North America. In the global status, 20 European species are threatened, susceptible, or rare.

On both continents, many genetically isolated stocks of salmon and trout confined to certain stream systems are threatened by eradication, or they have already been lost from a variety of threats such as river regulation and construction of dams that prevent upstream migration and that destroy spawning grounds. Discharge of wastewater from agriculture, industry, and towns has also contributed to the loss, as have acidification in areas with poorly buffered waters, inter-breeding with hatchery-reared individuals, and over-fishing in the ocean.

The Atlantic salmon once was very common along its range of distribution from Iceland to Portugal. It has now disappeared from many major rivers on the continent and the annual catch has dropped profoundly. The large salmon population in the Rhine supported an annual marketed catch of more than 100,000 individuals in the late 1800s. The species declined during the 1900s and went extinct in 1957, but was later reintroduced from an artificially reared stock. Atlantic salmon was also lost from all Danish rivers during the 1900s, apart from a small population (ca. 50 individuals) surviving in the Skjern River.

DNA studies of preserved salmon scales from the now extinct populations have shown that the individual rivers had genetically distinct populations. These differences, however, were much larger between populations in Denmark, Scotland, and Sweden than between populations in neighboring rivers. Thus, restocking with foreign salmon has been abandoned while attempts are made to secure and perhaps disperse the small, national salmon stock, which is probably better adapted to the local environmental conditions and food sources than are the foreign salmons. Although Atlantic salmon may still be part of the fauna in several countries of continental Europe, it has lost its role as an enjoyable catch and an important part of the ecology and food webs of the streams. Moreover, the original high genetic diversity among the many local populations of individual rivers is definitely gone.

B. Red Lists and Historical Development in Streams of Denmark and the United States

Red Lists focus on threatened, vulnerable, and rare species and usually include those species that have disappeared recently. Among the five studied groups of freshwater insects, the percentage of red-listed species of all national species in Denmark ranges from 32% for caddisflies to 50% for mayflies (Table III). Freshwater fish have 39% and amphibians have 36% of all national

TABLE III

Percentage of Species within Selected Groups of Freshwater or Mainly Terrestrial Animals on the Red List in North America (by 1990) and Denmark (by 1996)[a]

Group	North America	Denmark
Freshwater habitats		
Amphibians	28	36
Fish	34	39
Crayfish	65	—
Unionid mussels	73	—
Caddisflies	—	32
Dragonflies	—	42
Mayflies	—	50
Stone flies	—	40
Terrestrial habitats		
Birds	11	37
Mammals	13	30
Reptiles	14	25
Beetles	—	26
Butterflies	—	49
Moths	—	16

[a] Species on the Red List are recently extinct, threatened, vulnerable, or rare. Red Lists focus on particular vulnerable groups often highly valued by humans.

species on the Red List. Among terrestrial insects, the percentages vary from 16% for moths to 49% for butterflies. Birds and mammals have 37 and 30%, respectively.

Overall, the percentage of species on Red Lists is high for three reasons. First, the most vulnerable groups are more often included than the least vulnerable groups. Second, many species are naturally rare. Third, many species are present in very small numbers, because natural habitats have experienced profound areal restriction and deterioration. Thus, although the percentage of red-listed species tends to be higher for freshwater than for terrestrial species, this tendency disappears if more common groups of freshwater invertebrates such as dipterans, oligochaetes, and polychaetes were included.

Consequently, Red Lists have several weaknesses because they are selective and qualitative, and the intensity of search for rare species has increased over time. Red Lists do not give quantitative data on the abundance of species, and they can only describe the temporal development in crude ways provided evaluations are repeated at suitable intervals.

Many evaluations of species development, however, suffer from the lack of suitable historical description of species distribution and abundance. If such studies indeed exist, they did not use exactly the same methods and survey intensity and are therefore open to critique, even though differences may be very profound and without reasonable doubt are real. An analysis of *Potamogeton* (a large, aquatic plant genus) in 13 localities in Danish lowland streams, for example, revealed the existence of 6.0 species per locality and 16 species altogether 100 years ago; today the mean number is 2.8 species per locality and only 7 species grow in the same 13 localities. Several of those species that have disappeared from the 13 localities have become extremely rare throughout the country. Overall, the vegetation has become poorer in species and ecological types. Mainly oligotrophic or large, slow-growing species have disappeared, and the few survivors are species of high dispersal capacity and tolerance to disturbance and eutrophication (e.g., *P. crispus* and *P. pectinatus*); these species have now established a more profound dominance. The same type of development has taken place in other lowland regions of northwestern Europe such that some species (e.g., *P. acutifolius*, *P. filiformis*, *P. zosterifolius*) have become rare or threatened over wide areas.

It is noteworthy that plant species in streams display an overall positive relationship between geographical range size and local abundance, resembling the pattern more thoroughly described for terrestrial herbs and animals. There is an overall transition from species of low geographic range size and low local abundance to species of high range size and high local abundance. Those first-mentioned species face a double jeopardy of extinction because they grow in just a few places and they are infrequent at sites where they do occur. Stochastic loss of some habitats and degradation of others should therefore have a strong impact on these species, because their few and small populations make them susceptible to further losses and reduce their ability to disperse to new suitable places that may arise. In contrast, widespread species of high abundance have a double security, because they have a higher probability of surviving stochastic changes and spreading to new habitats.

Freshwater insects have undergone a similar decline of species richness, number of ecological types, and number of occupied habitats. About 20 species of the 285 Danish species of caddisflies, dragonflies, mayflies, and stone flies have gone extinct during the last 100 years. An even large number of 76 species are either threatened, vulnerable, or rare. The red-listed species often require high water quality and have long life cycles that are sensitive to disturbance. The historic development has, therefore, led to a more stereotypic composition of both plant and macroinvertebrate communities with the same few robust species dominating at most stream sites. Species holding refugia within the stream systems are likely to recover within a few years following environmental improvement via redistribution by upstream and downstream migration. However, recovery is slow for species that have been lost from entire stream systems. Recovery may take from decades to centuries, or not occur at all, if species have been lost from islands or from geographically isolated regions remote from possible founder populations.

There are very few similar historical evaluations of species distribution and abundance. What come closest are studies of agricultural areas of the U.S. Midwest. These demonstrate profound reduction of habitat quality, species richness, and abundance of stream invertebrates and fish over the last 50 years as a result of removal of most of the natural riparian forest vegetation, stream regulation, and alteration of water chemistry. Indices of stream quality decline with the intensity of cultivation in the catchment, and this is accompanied by a predictable decline in species richness and feeding types of invertebrates and fish. Predominantly robust, generalist species have survived in the stream communities. Fish specialized on insects have disappeared, while omnivorous fish have survived. The historical development of plants and animals, therefore, has many similarities in the lowland streams of Europe and the United States, because all macroscopic organisms have been strongly influenced by cultivation and agricultural practices.

C. Biological Consequences of Organic Pollution and Agricultural and Industrial Chemicals

Organic pollutants give rise to deoxygenation and release of ammonia and other mineral nutrients in the water due to high microbial activity. Organic pollution leads to unhealthy aquatic conditions, high nutrient levels, untransparent water, and smothering of sediments. These conditions are detrimental to the survival of invertebrates and fish, and highly oxygen-sensitive groups of insects such as beetles, caddisflies, mayflies, and stone flies, as well as salmonid fish, may disappear almost entirely. Also, the diversity of feeding groups of macroinvertebrates declines as stony substrata for algal grazers are smothered, coarse terrestrial detritus for shredders becomes insignificant relative to the accumulation of organic mud for deposit feeders, and large predators with long life cycles find it hard to sustain episodes of oxygen depletion. Thus organic pollution is accompanied by a general loss of species richness, taxonomic diversity, and functional diversity of macroinvertebrates, in addition to the appearance and dominance of particularly robust species (e.g., tubificid polychaetes, *Chironomus* and *Psychoda*) in exchange for the loss of many more sensitive species (Fig. 3).

The biological changes are principally the same in temperate and tropical streams. The macroinvertebrate indices originally developed to describe organic pollution in temperate streams work well in tropical streams. The obvious reason is that the taxonomic composition of orders and families of macroinvertebrates does not differ greatly between temperate and tropical streams. Taxonomic similarity is particularly high between temperate lowlands and tropical highlands. The warm tropical lowland streams may have a high sensitivity to organic pollution because of low oxygen solubility, fast microbial degradation, and high invertebrate metabolism. Tropical highland streams also may be more sensitive to organic pollution than temperate streams because of the reduced partial pressure of oxygen at higher elevations, while temperatures resemble those in temperate lowland streams.

Domestic wastewater contains many chemical stressors in addition to the oxygen-consuming organic material. Sensitivity of organisms to low oxygen often goes hand in hand with the sensitivity to industrial chemicals. Similarly, sensitivity to acid waters usually changes in parallel with the sensitivity to heavy metals. Chemical stressors are lethal at high concentrations, and sublethal disturbance of body form, reproduction, neural function, and behavior occurs at lower concentrations. Among *Chironomus*, for example, the proportion of individuals with deformed mouthparts increases with the level of pollution. Among net-spinning caddis larvae, the proportion of abnormal, incomplete nets increases with the concentration of toxic chemicals such as aluminum. These sublethal effects will eventually lead to a decline of populations due to reduced growth and survival compared with more resistant species. However, under very severe oxygen deficiency or high concentrations of toxic chemicals, no species of higher life-forms among fish, invertebrates, and plants are capable of surviving, and only microorganisms will remain. Chemically polluted streams and ponds without any higher life include habitats with high concentrations of cadmium, copper, zinc, and ferrous iron.

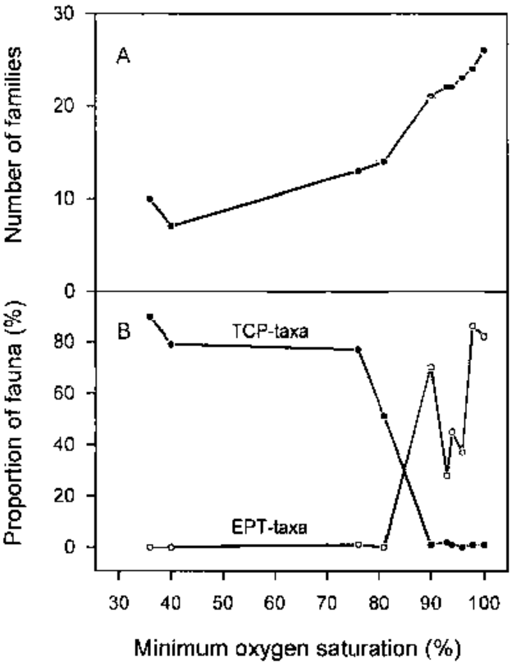


FIGURE 3 Composition of macroinvertebrate communities in small highland streams in tropical Ecuador as a function of organic pollution from villages, resulting in a lower minimum oxygen saturation in the water. (A) Number of macroinvertebrate families. (B) Proportion of oxygen-requiring taxa of caddisflies, mayflies, and stone flies (EPT taxa) to taxa that require less oxygen (TCP taxa), such as Tubificidae, *Chironomus*, and *Psychoda*. (Data from Jacobsen, 1996.)

The immune and hormone systems of animals can be disturbed by exposure to organochlorine compounds (e.g., DDT, PCBs), nonylphenols and octylphenols of industrial detergents, and an estradiol component from degradation of contraceptive pills. In water bodies with a high proportion of sewage effluents, such compounds can apparently lead to the development of hermaphrodite individuals of roach and rainbow trout, abnormally small testicles in alligators, and feminization of populations of terns and gulls feeding on aquatic animals. Estrogenic effects have received great attention because human populations have been showing increasing frequencies of testicle cancer and decreasing numbers of viable sperm cells over the last 50 years in Scandinavia.

The great variety of pesticides now being found in surface waters are also potentially dangerous because they have been manufactured to be deadly poisonous in trace amounts to fungi, insects, and plants. Insecticides target the most species-rich group of animals on land and in freshwater. For the time being, however, the broad-scale environmental significance of pesticides is largely unknown. There are suggestions that intense application of pesticides in some regions can account for an impoverished insect fauna in streams, where other pollution sources have largely been eliminated.

D. Biological Consequences of Lake Eutrophication

Lake eutrophication influences ecosystem structure and functioning as well as the distribution and abundance of most species. This fundamental influence is the result of the primary role of nutrients in boosting the biomass and productivity of phytoplankton, setting the scene for many structural and functional regulations.

The attitude toward cultural eutrophication changes with its magnitude and the public interests in different regions. In lowland regions of Europe and North America, substantial eutrophication has shifted the natural wide range of lakes to a higher nutrient level that encompasses conditions from mesotrophy to hypereutrophy. Water purification aims at improving hypereutrophic lakes and moving them to eutrophy, such that most lakes may occupy the narrow range from mesotrophy to eutrophy with no lake being unpolluted and no lake being extremely polluted. As a consequence, species assemblages have become more stereotypic as the susceptible species of oligotrophic habitats, and in some cases also the extremely robust species of hypereutrophic habitats, have become very rare. Full control of pollution and the ecological disturbance associated with agriculture, fish farming, forestry, and recreational use, which may be needed to maintain pristine, oligotrophic, clear-water lakes, has rarely been implemented.

Social and environmental priorities are different in other societies. In Southeast Asia, for example, the focus is more on fish production and less on aesthetics and the protection of rare species. A small lake that is heavily fertilized by village sewage may be a valuable source of fish that feed on the thick algal soup and that are tolerant to periods of deoxygenation (Moss, 1998).

1. The Case of Lake Fure and Other Hardwater Lakes

The 100-year development of Lake Fure in Denmark can serve as an example of the widespread, accelerated nutrient loading of temperate lakes followed by recent reduction of loading rates (see Fig. 1). The long-term evaluation for this lake matches the long water residence time (15 years) and the slow numerical response of macrophyte communities in this naturally mesotrophic, hardwater lake.

Annual external phosphorus loading of the lake increased from 1.3 ton of P in 1900, to over 10 tons in 1950, to 37 tons in 1969. Until the late 1940s, the lake had managed to tolerate the increased loading relatively well because of the efficient sediment binding of phosphorus to calcium carbonates and iron oxy-hydroxides. The bottom waters had remained oxic and iron-bound phosphorus had stayed in the sediments. With the accelerating phosphorus load during the 1950s and 1960s, extensive phytoplankton blooms developed, leading to oxygen-free bottom waters and release of iron-bound sediment phosphorus. Following diversion and tertiary treatment of the sewage in 1969, the external phosphorus loading declined to 3–4 tons P, but the internal concentrations have remained high owing to sediment release of about 15 tons P year^{-1} from the large P-pools that have accumulated.

The mean Secchi-depth during summer was about 5–6 m from 1900 to 1940, but it has fluctuated between 1.5 and 3.2 m since the early 1970s. The number of phytoplankton species has remained approximately constant between 70 and 85. However, species within particularly nutrient-demanding groups typical of eutrophic lakes (e.g., blue-green and chlorococcalean green algae) have increased from 17 to 39, and species typical of oligotrophic lakes (e.g., chrysophytes, desmids, and certain dinoflagellates) have declined from 32 to 13. Certain species of blue-greens, chrysophytes, and dinoflagellates, which were able to live in the deeper and colder waters within or below the thermocline at 10–16 m depth, have disappeared with the restriction of the illuminated zone to the upper few meters of the water column.

Species richness of submerged macrophytes in Lake Fure before the recent cultural eutrophication commenced was 35 species, including 18 rooted flowering plants, 10 characeans, and 5 mosses. Tall flowering plants grew from shallow water to the depth limits at 8 m, but dominated the vegetation down to 5 m depth. Small flowering plants, characeans, and mosses formed a mixed carpet of vegetation below the canopy of large flowering plants, while characeans dominated from 5 to 8 m.

Between 1983 and 1993 only 10 of the original 35 species remained, whereas 3 new species of pollution-tolerant green macroalgae had appeared. Small flowering plants and all mosses and characeans had vanished. Eight large species of canopy-forming flowering plants have survived in the turbid lake, probably because they can compensate for the poor light conditions by exposing the leaves close to the water surface. Depth distribution of all original submerged macrophyte species has been restricted from 8 m to 2–4 m because of the reduced light penetration.

The development reveals a reduction of species richness, taxonomic diversity, and life-forms of submerged macrophytes. In addition, there is now a greater dominance among the few surviving species in the present turbid phase of the lake. Certain species have disappeared because of the restriction of suitable growth habitats, and other species have been outcompeted by the canopy-forming plants and the fast-growing macroalgae that proliferate at higher nutrient availability in water and sediments. The plants increasing in rank among the surviving species in Lake Fure are the same eutrophic species (e.g., *Potamogeton crispus* and *P. pectinatus*) that have become more abundant in lowland lakes and streams throughout northwestern Europe.

The steep decline of aquatic mosses and characeans in Lake Fure is symptomatic of the broad-scale changes in eutrophicated temperate lakes. Characeans include about 80 species globally. In countries experiencing profound lake eutrophication, the characean group has become rare and many species have been added to the Red Lists. Formerly oligotrophic, hard-water lakes and brackish lagoons dominated by species-rich communities of characeans now represent a threatened habitat that is included in the European Community Habitat Directive.

2. Broad-Scale Comparisons and Whole-Lake Experiments

Other individual lake studies basically tell the same history of eutrophication. In the large lakes, however, the role of the littoral zone and bottom plants is small, and emphasis is greater on the increase of phytoplankton biomass, loss of bottom fauna, and alteration of fish communities. In most lakes, phosphorus is the main nutrient limiting phytoplankton biomass, but in some lakes the importance of phosphorus and nitrogen as limiting nutrients alternates with the seasons, and in other lakes nitrogen plays the major role (e.g., Lake Tahoe on the California–Nevada border).

In lakes with a small magnitude or a short period of cultural eutrophication, rapid recovery is observed upon reduction of nutrient loading, because only small nutrient pools have accumulated in the lake bottom. Recovery also is fast and more complete when water renewal is high, and when large proportions of unpolluted water are available from undeveloped mountain and forest areas such that nutrients can be flushed from the lake (e.g., Lake Washington in the state of Washington).

Case studies of individual lakes, comparisons among many lakes, and controlled whole-lake fertilization experiments all show strong positive relationships among nutrient availability, phytoplankton biomass and productivity, zooplankton and fish production, light attenuation in the water column, and risk of anoxia in bottom waters. Predictions of phytoplankton biomass based on external phosphorus loading and water renewal are often excellent for lakes within a region of the same climate and soil geology and for lakes where nutrient inputs are precisely known. These results suggest that biotic differences do not play a major role for the biomass of phytoplankton. It has therefore been argued that the lower accuracy of predictions observed among lakes from different regions and studies may result from methodological differences and uncertainties rather than from the influence of biological differences in the food webs. In shallow lakes, however, wide shifts sometimes take place from transparent lakes of low phytoplankton biomass and high macrophyte cover to more turbid lakes of higher phytoplankton biomass and marginal macrophyte cover, even though the rate of phosphorus loading remains approximately the same. In shallow lakes, rooted macrophytes and attached animals have the potential to exert considerable control on the communities in the open waters—in part through internal alterations of nutrient cycling. The same intense interactions between organisms and processes at the bottom and in the open water are not possible in large, deep lakes.

3. Resource and Predatory Control in Food Webs

In many lakes, phosphorus stripping and the diversion of sewage have reduced nutrient input, but improvement of lake quality has been minimal because of high internal nutrient circulation. To reduce sediment release and internal water concentrations of phosphorus, several methods have been attempted. One principle is to remove P-enriched cultural sediments by pumping them up. Another principle is to increase the P-binding capacity in the sediment by adding iron and ensuring oxidized conditions conducive to efficient phosphorus binding to iron oxy-hydroxides by aerating the bottom waters or injecting nitrate.

In shallow lakes, biomanipulation has had the objective of reducing the number of fish, which eat the large zooplankton (especially daphnids), and thereby hopefully reduce the biomass of phytoplankton through enhanced zooplankton grazing. The additional goal is to increase the cover of submerged vegetation, which can directly reduce sediment release of nutrients, impede phytoplankton development by shading, and provide refuge to daphnids.

The success of attempts to reduce phytoplankton blooms and improve water clarity by internal manipulation of sediments and food webs is variable, and some positive short-term effects have not been sustainable in the long term. Opinions on the statistics of successes and failures of biomanipulation often vary along with the subjective view on the relative importance of resources and predatory control in the food webs. The view that grazing control on phytoplankton is high in lakes with an even number of trophic levels in the food web and low in lakes with an odd number of levels has received some support, but also attracted growing opposition. Among reasons for the opposition are (1) predictions based on the concept are poor, (2) it is difficult to define trophic levels and numbers of links, and (3) many species are omnivorous, show ontogenetic changes in food preference, and cover more than one trophic level. Moreover, there is a strong positive relationship from nutrients and resource levels to biomass and productivity of phytoplankton, zooplankton, and fish because of the need to transfer energy and matter from lower to higher trophic levels.

Many studies clearly demonstrate a strong role of predators on size composition and species abundance of prey organisms, and they show behavioral alterations of predators and prey to increase their own growth and survival. These flexible biological responses, however, do not imply that the main control of the combined biomass and productivity of all the species at each trophic level is exerted by predation. Overall, the predatory fish have a small cascading trophic effect through the entire food web and a minor influence on phytoplankton biomass and water transparency. As a consequence, there is no easy technical or biological fix to ameliorate lake eutrophication, but various techniques may assist the effect of sustained reduction of external nutrient input with the aim of reducing unwanted phytoplankton blooms.

4. Arctic and Tropical Lakes

Consequences of eutrophication are also strong in Arctic and tropical lakes. Although Arctic lakes are ice-covered for most of the year and have very low temperatures, nutrient limitation is a main determinant of the variability in phytoplankton biomass and productivity. Arctic lakes, being surrounded by frozen catchments with low rates of mineralization, tend to be more nutrient-poor and have lower rates of phytoplankton photosynthesis per unit of chlorophyll pigment than Arctic marine waters. Temperatures close to zero reduce enzyme activities, and an efficient way for algae to compensate is to achieve higher enzyme concentrations by greater nitrogen investment.

Arctic lakes are often shallow and covered by ice and snow for extended periods. They are, therefore, highly sensitive to oxygen depletion following degradation of a nutrient-stimulated phytoplankton productivity. No replenishment of oxygen is possible at the ice-covered surface, no oxygen is produced by photosynthesis owing to lack of light for many months, and the initial oxygen pool is limited by shallow depths. The presence of fish in Arctic lakes is critically dependent on water depths and the availability of oxygen throughout the year.

Heavy use of fertilizers is spreading in many tropical regions with the growth in intensive farming, and expanding tropical cities are releasing increasing amounts of untreated sewage. Lake Victoria in East Africa illustrates the catastrophic increase in this problem. This huge, ancient lake used to contain many unique species of haplochromine fish, and the fishery supported the protein needs of many people. From 1960 to 1990, the phytoplankton community showed a 10-fold increase in biomass and a shift from the dominance of diatoms to troublesome blue-green algae. Earlier, oxygen depletion was restricted to limited areas just above the deepest sediments, but nowadays oxygen-free conditions and fish-kills are widespread in the bottom waters. The most likely explanation for these changes is the clearing of woodlands, with the subsequent increase in erosion and inflow of sediment, and the fast rise in human population density, resulting in the release of larger volumes of sewage. The introduction of Nile perch has completely altered the food webs, reduced the important stocks of tilapias, and apparently driven many of the small haplochromine fish species to extinction.

In the 1990s, there has been a prolific spread of water hyacinth along the coasts of Lake Victoria, leading to severe problems for the fishery and boat traffic. The mass growth of floating-leaved plants in lakes, rivers, and canals is a spreading problem in subtropical and tropical regions, and it is enhanced by eutrophication and introduction of exotic pest species such as *Azolla*, *Salvinia*, and water hyacinth, which lack their natural competitors, herbivores, and pathogens in these new habitats.

5. Species Richness and Lake Trophy

Species richness increases with habitat area and heterogeneity. The increase of species richness (S) with habitat area (A) is often predicted by the equation $S = \text{constant} \times A^z$. The z-value depends on the organisms, the spatial range of the study, and the habitats, and there is no single mechanism that can account for the variability of z. However, low z-values are consistent with high immigration rates, low extinction rates, and a low rate of increase in additional habitat with increasing area. As a consequence, very small z-values are predicted for bacteria, microalgae, protozoans, and zooplankton. Indeed, the z-values predicting crustacean zooplankton species richness are only 0.054 for European and 0.094 for North American lakes. A low z-value of 0.10 for submerged macrophytes in Scandinavian lakes is also consistent with their widespread occurrence.

In contrast, species richness of fish in lakes of four different regions in Canada and the United States shows much higher z-values (0.16, 0.22, 0.36, and 0.37) consistent with the more restricted geographic range size, smaller local abundance, and greater risk of extinction of fish following disturbance or restriction of natural habitats. As a consequence, risks of local, regional, and global extinction are high among fish and low among microorganisms and plankton organisms, with aquatic plants and insects probably holding an intermediate position. Some large crustaceans and molluscs presumably resemble the fish by having a low rate of dispersal and including several species of restricted geographical distribution (e.g., the freshwater crabs and prawns on tropical oceanic islands).

Changes in ionic composition and nutrient concentrations of freshwaters are superimposed on changes in habitat area and heterogeneity. Water chemistry is not independent of habitat area and heterogeneity, since small lakes include a higher proportion of low-pH, low-calcium, and nutrient-poor waters than do large lakes. In many studies of the influence of pH and lake trophy on species richness, it is not possible to fully compensate for the influence of habitat area and heterogeneity.

The most diverse conditions and highest species richness often are found under mesotrophic conditions, where many species can co-exist at different sites and depths within the lakes. Thus, high habitat heterogeneity is probably important for the common peak of species richness under mesotrophic conditions. For example, many free-living and attached species adapted to different conditions of light, temperature, and exposure can replace each other along depth gradients from shallow to deep waters. Under very oligotrophic conditions, many algae and plant species are nutrient-limited, and low primary production places an energetic restriction on the diversity and density of invertebrates and fish. In contrast, under hypereutrophy, many species are restricted by the lack of light and oxygen in deep waters, and special adaptations are required for survival in very muddy sediments.

Two other conditions contribute to the decline of species diversity under hypereutrophic conditions. First, the natural local variability between unfertilized and fertilized sites disappears with the overall nutrient enrichment at all sites accompanying eutrophication. Second, hypereutrophy is exceptional for most species, which have evolved over many millions of years of no or weak human impact. Species richness is expected to be highest in those common habitat types that have had the most widespread and long-term natural occurrence, because there has been time and room for speciation. Hypereutrophic lakes, which are common today, have been exceedingly rare during most of the development of freshwater species.

It is noteworthy that lake ecosystems are much more sensitive to environmental deterioration and catastrophic declines of organisms by eutrophication than terrestrial ecosystems for at least four reasons. First, nutrients tend to stay in circulation within the lake boundaries. Second, phytoplankton responds by a steep increase in biomass and productivity. Third, photosynthesis and growth of algae and plants are restricted to the uppermost surface waters because of the impoverished light conditions. Fourth, the risk of oxygen depletion is very high because of low oxygen solubility in the water. Therefore, species richness within many groups of organisms declines in hypereutrophic lakes because of reduced habitat heterogeneity, restriction to the distribution of organisms, and development of stressful and highly variable environmental conditions with respect to light, oxygen, pH, and sulfide.

E. Biological Consequences of Acidification

Acidification initiates complex chemical and biological changes in surface waters and catchments with poorly weathered rocks and thin soils. The acidification of freshwater habitats is also accompanied by acidification of terrestrial ecosystems in the catchments, leading to a widespread regional decline of biodiversity.

Whole-lake experiments involving acidification to pH 5 demonstrate that damage to ecosystem functioning is of secondary concern compared to impoverishment of the biotic communities. Although many phytoplankton algae disappear, the remaining algae can maintain biomass and primary production at the same level, largely regulated by the input of phosphorus and water renewal. The rate of decomposition is governed by primary production and not by pollutants. However, the complete disappearance of large invertebrate detritivores upon acidification may lead to the accumulation of detritus and over the long term reduce plant production based on recirculated nutrients. Overall, nutrient cycling was not disrupted in whole-lake experiments, but nitrification declined at pH values below 5.4.

Acidification of the watershed changes ionic composition and may reduce availability of phosphorus in soils and surface waters. Soil acidification increases the leaching of base cations, and it has specific effects on the solubility of aluminum and phosphate. Chemical weathering of aluminum minerals is enhanced under acid conditions. However, soluble forms of aluminum ions may precipitate with phosphate at the proper pH, leading to "oligotrophication." As a result, lakes in acidified watersheds may become very clear from the enhanced phosphorus limitation of phytoplankton growth, and certain acid-tolerant aquatic mosses (e.g., *Sphagnum* spp.) and filamentous green algae (e.g., *Mougeotia* spp.) may spread across the lake bottom to greater depths. Only a few vascular plant species can tolerate highly acidic conditions (e.g, *Juncus bulbosus*).

Species diversity and abundance decline within most groups of organisms upon acidification, though some groups are much more sensitive than others and distinct differences exist among species in their tolerance to acidification (Fig. 4). In acidified regions of Europe and the U.S. Northeast, many lakes have lost 30% or more of the species in some taxonomic groups. Among planktonic and benthic microalgae, desmids—a group of green algae—prefer buffer-poor waters of low pH and low calcium content, and they are often accompanied by a few acid-tolerant species of blue-green algae, dinoflagellates, diatoms (e.g., *Eunotia exigua*), green algae (*Chlamydomonas acidophila*), and euglenophytes (*Euglena mutabilis*). In contrast, most species of diatoms, chlorococcalean green algae, and blue-green algae prefer well-buffered waters of neutral pH. Among planktonic rotifers, copepods, and daphnids, only a few species tolerate acid waters (<4–5), so the planktonic food webs become simple. The tolerant species can utilize the vacant ecological niches left by others, thus maintaining a relatively constant zooplankton biomass despite the decline of species richness.

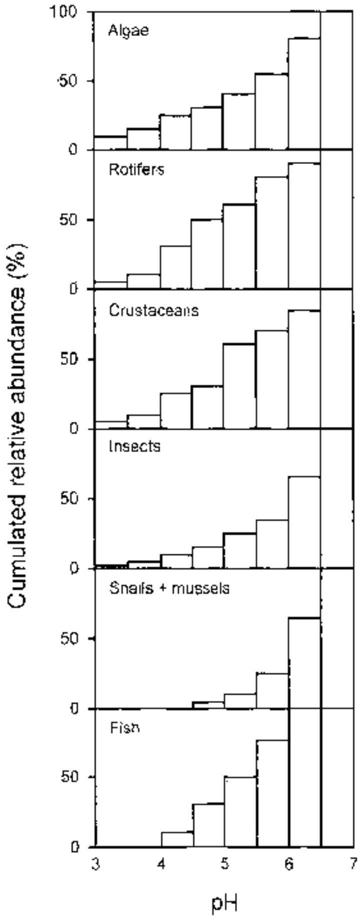


FIGURE 4 Cumulative frequency distribution of minimum fields observations of pH for 46–370 aquatic taxa within different taxonomic groups. (Data compiled by Mills and Schindler, 1986, adapted from Schindler, 1991.)

Many snails, mussels, and crustaceans (e.g., crayfish) are highly sensitive to acidification due to problems in maintaining the ionic regulation of sodium, chloride, and potassium and in forming the exoskeleton of calcium carbonate or calcium-chitin when pH and calcium concentrations are low. Higher calcium concentrations can help in maintaining membrane integrity, reducing gill permeability to salt loss, and ensuring the construction of shells. Several metal toxins such as aluminum mobilized under acidic conditions have the opposite influence.

Among insects, the most acid-tolerant species include surface-breeding beetles with a hard exoskeleton and certain stone flies and caddisflies with a low metabolism and mostly lacking thin-walled body parts. In contrast, mayflies with their large, thin gills are often highly sensitive to acidification. The decline of species richness of insects is not as pronounced when acidic conditions occur in waters rich in humic substances, which can bind and detoxify the metal ions mobilized at low soil pH.

A common result of acidification is the dramatic loss of fish. Acid-stressed regions of Scandinavia, the Adirondack Mountains and Wisconsin (both in the United States) and La Cloche Mountains (Canada) reveal a steep decline of fish species as lake pH declines from 8 to 4, and many lakes with pH less than 5 are fishless. The decline of fish species richness with reduced pH is much less pronounced in subtropical Florida lakes and in tropical lakes. The presence of more acid-tolerant fish species in warm lakes may be due to a much slower and more gradual decline of pH over time and, accordingly, better opportunities of genetic adaptations to low-pH waters. Also, Florida lakes have lower concentrations of aluminum and free mineral acids, longer growing seasons, and no episodic input of acids resembling the massive acid input of north temperate lakes during snowmelt.

Fish studies in north temperate lakes show that eggs, embryos, and larval stages of fish are more sensitive to acidification than the adult stages. As a consequence, young generations may vanish upon acidification, whereas old fish may remain for a while. Populations will eventually die out unless young fish are continuously stocked. Salmonids and cyprinids with a high metabolism and a large gill surface appear to be more susceptible to acidification than fish with a low metabolism and a low gill surface (e.g., eels). A critical phase for the well-being of fish and invertebrates is the maintenance of efficient oxygen uptake and ion regulation across the gill surfaces.

1. Ecosystem Functions and Explanations

Ecosystem functions have proven to be more resistant to acidification than species changes. Species assemblages are often plastic enough to maintain ecosystem functions under quite extreme stresses, although only a few species survive. For example, plant productivity continued in a dense monospecific vegetation of *Juncus bulbosus* growing in a highly acidic, temperate stream (pH 2–3) from a lignite mining area with intense pyrite oxidation, while neighboring alkaline streams (pH 6–7.5) contained a mixed assemblage of about 10 rooted plant species. Similarly, phytoplankton biomass and production were unaffected in lakes polluted with strong acids and trace metals from smelters. Phytoplankton and other microorganisms are also locally abundant and are easily dispersed, such that highly pH-tolerant species will rapidly show up and multiply in recently acidified localities.

Some ecosystem processes, however, may decline or disappear entirely under very acidic conditions. At pH values below 5.4–5.7, ammonium begins to accumulate due to the cessation of nitrification. Therefore, the loss of internally produced nitrate through denitrification will stop. However, the enhanced atmospheric deposition of nitric acid and sulfuric acid will permit the microbial activity of denitrifiers and sulfate reducers in oxygen-free sediments and, thereby, generate alkalinity in the water bodies.

The biotic communities of lakes and streams are strongly damaged and impoverished by acidification. The most vulnerable species appear to be organisms with year-long life cycles and poor dispersal, such as large invertebrates and fish. These groups are also the less speciose, particularly in northern water bodies, which means that there is little redundancy in the food webs to prevent pH-sensitive species of large detritivores, plankton-eating fish, and predatory fish from being replaced by pH-tolerant species. Fish and crayfish, which are highly valued by humans, are therefore easily damaged. Even bird populations such as dippers, osprey, and pied flycatchers may vanish from the lack of food or the formation of thin-shelled eggs while feeding on calcium-deficient and aluminum-rich prey.

Overall, surface waters of very low pH represent a special environment to which very few species have become adapted as a result of (1) the direct stress at low pH, (2) the unsuitable ionic composition, (3) the toxic metals, and (4) the rarity of these habitats during the evolution of aquatic organisms, particularly in temperate and subarctic regions.

VI. CONCLUSIONS

Freshwaters of high quality are essential to humans for domestic and agricultural uses. Freshwater ecosystems provide many other highly valued and critical ecological services, and they support a very high diversity of species and taxonomic groups considering their relatively small volumes and surface areas.

Human impact has caused numerous stream reaches and shallow lakes to disappear, disturbed and polluted many others, and disrupted the close contact between the floodplain and the streams. Consequently, the diversity and abundance of numerous freshwater species have declined due to the profound restriction of areal cover and habitat diversity of freshwater environments. Moreover, the systematic loss of pristine, unpolluted freshwaters has threatened or eradicated those species that require such environments.

The overall disruption of the linkage between freshwater and terrestrial environments has far-reaching consequences for global hydrology, elemental cycling, and biodiversity. Among those consequences are enhanced risks of flooding and drought, greater risks of land erosion and marine siltation, and reduced opportunities for the countless plants and animals who live a natural double-life between freshwater and land to survive in the future.

Accompanying all of this disturbance and pollution has been the accidental or deliberate introduction of nonnative organisms such as crayfish, plants, sport fish, and zebra mussels, which may monopolize freshwater habitats and drive the native, endemic species in formerly undisturbed lakes and river systems to extinction. Several species of freshwater fish have gone extinct during the last few centuries, and many additional species of plants, large invertebrates, fish, and frogs now exist in extremely small and geographically isolated populations. These species face a great risk of extinction, and they can no longer provide their accustomed ecosystem functions. Nor can these species be enjoyed within most of their native area. Unfortunately, they will find it exceedingly difficult to recolonize newly restored aquatic habitat and contribute to future speciation.

See Also the Following Articles

AGRICULTURE, INDUSTRIAL • ENDANGERED FRESHWATER INVERTEBRATES • EUTROPHICATION/OLIGOTROPHICATION • FRESHWATER ECOSYSTEMS • HUMAN EFFECTS ON ECOSYSTEMS, OVERVIEW • POLLUTION, OVERVIEW • WETLANDS RESTORATION

Bibliography

Allan, J. D. (1995). *Stream Ecology. Structure and Function of Running Waters.* Chapman & Hall, London.

Allan, J. D., and Flecker, A. S. (1993). Biodiversity conservation in running waters. *BioScience* 43, 32–43.

Jacobsen, D. (1996). *Ecology and Environmental Status of Tropical Streams in Ecuador*, Freshwater Biological Laboratory and Danida, Danish Foreign Ministry, Copenhagen.

Kristensen, P., and Hansen, H. O. (eds.). (1994). *European Rivers and Lakes—Assessment of Their Environmental State*, EEA Environmental Monographs 1. European Environment Agency Copenhagen.

Moss, B. (1998). *Ecology of Fresh Waters. Man and Medium, Past to Future*, 3rd ed. Blackwell, Oxford, United Kingdom.

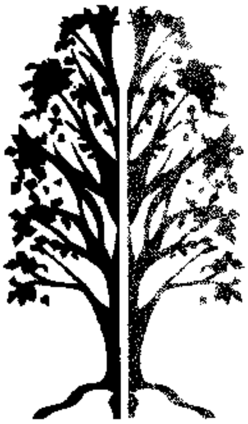
Palmer, M. A., *et al.* (1997). Biodiversity and ecosystem processes in fresh-water sediments. *Ambio* 28, 571–577.

Petts, G. E. (1984). *Impounded Rivers.* John Wiley & Sons, Chichester, United Kingdom.

Sand-Jensen, K., and Pedersen, O. (eds.). (1997). *Freshwater Biology. Priorities and Development in Danish Research.* G.E.C. Gad, Copenhagen.

Schindler, D. W. (1988). Experimental studies of chemical stressors on whole lake ecosystems. *Verh. Int. Verein. Limnol.* 23 11–41.

Schindler, D. W. (1991). Aquatic ecosystems and global ecology. In *Fundamentals of Aquatic Ecology*, 2nd. ed. (R. S. K. Barnes and K. H. Mann, eds.), pp. 108–122. Blackwell, Oxford, United Kingdom.



FUNCTIONAL DIVERSITY

David Tilman
University of Minnesota, St. Paul

GLOSSARY

diversity-productivity hypothesis The proposal that greater diversity would lead, on average, to greater total biomass or productivity.

diversity-stability hypothesis The proposal that ecosystems containing more species would be more stable.

ecosystem composition The list of species or functional groups that are present in a given ecosystem.

ecosystem functioning The rate, level, or temporal dynamics of one or more ecosystem processes such as primary production, total plant biomass, or nutrient gain, loss, or concentration.

functional diversity The range and value of those species and organismal traits that influence ecosystem functioning.

functional group A set of species that have similar traits and that thus are likely to be similar in their effects on ecosystem functioning.

niche differentiation Differences in the morphology, physiology, or behavior of species that can influence their abundances, dynamics, and interactions with other species, including the ability of various competing species to coexist.

sampling effect The hypothesis that diversity might influence an ecosystem process because of the greater chance that a given species trait would be present at higher diversity, and the effect of its presence on ecosystem functioning.

FUNCTIONAL DIVERSITY refers to those components of biodiversity that influence how an ecosystem operates or functions. The biological diversity, or biodiversity, of a habitat is much broader and includes all the species living in a site, all of the genotypic and phenotypic variation within each species, and all the spatial and temporal variability in the communities and ecosystems that these species form. Functional diversity, which is a subset of this, is measured by the values and range in the values, for the species present in an ecosystem, of those organismal traits that influence one or more aspects of the functioning of an ecosystem. Functional diversity is of ecological importance because it, by definition, is the component of diversity that influences ecosystem dynamics, stability, productivity, nutrient balance, and other aspects of ecosystem functioning.

Encyclopedia of Biodiversity, Volume 3

I. MEASUREMENT OF FUNCTIONAL DIVERSITY

Because of the large number of traits that each species possesses, the large number of different species that exist in most habitats, and the incomplete knowledge of which species traits influence various ecosystem processes, there is, as yet, no simple way to measure functional diversity. Rather, items that are more easily measured than functional diversity are used as indices or correlates of functional diversity. The most common of these indices is the number of species present in a habitat, which is called the species richness or species number of the habitat. All else being equal, habitats with greater species richness should also have greater functional diversity. This occurs because species differ in their traits. Sites that contain more species should thus also contain, on average, a greater range of species traits, which is greater functional diversity. Species diversity indices, such as the Shannon diversity index, are similarly used as indirect measures of functional diversity. Another commonly used index of functional diversity is the number of different functional groups (defined later) that exist within a given community or ecosystem. This is also called functional group diversity. Assuming that organisms can be categorized as belonging to groups that differ in traits relevant to ecosystem functioning, greater functional group diversity should correlate with greater functional diversity. However, variations among species within a given group could also contribute to functional diversity. Observational, experimental, and theoretical studies indicate that functional diversity, as measured by any of these three means, is one of several important factors that determine ecosystem functioning. Because there is, as yet, no clear way to measure functional diversity, one or more of these three indices will be used as a proxy for functional diversity in this chapter. Before reviewing the research linking functional diversity to ecosystem processes, which is the focus of the remainder of this chapter, it is important to introduce and define some terms.

II. EXPLANATION OF CONCEPTS AND TERMINOLOGY

A. Functioning

As they are used by ecologists, the words *function*, *functional*, and *functioning* are not meant to imply that an ecosystem process has any underlying goal or purpose. Indeed, to try to minimize any such implications, it has become standard practice to refer to "ecosystem functioning" or "ecosystem process" rather than the "function of an ecosystem." The latter might be misinterpreted as meaning that an ecosystem exists to perform a given function, which is inconsistent with our knowledge of the process of evolution. Rather, functioning refers solely to the way in which an ecosystem operates.

B. Ecosystem Processes

Ecologists study many different aspects of the functioning of communities and ecosystems. The three most frequently considered ecosystem processes are productivity, stability, and resource dynamics. Productivity refers to the rate of production of biomass within a given trophic level. The production of plant biomass is called primary production, the production of biomass of herbivores is called secondary production, and that of predators is called tertiary production. Stability has a wide range of definitions, including the degree to which an item is resistant to change when experiencing a single perturbation, the degree to which an item fluctuates in response to an ongoing suite of small-scale perturbations, and the dynamics of return to its prior state after a single perturbation. Stability can be measured at the level of populations, communities, or ecosystems. The resource dynamics of an ecosystem are measured by the rates of supply and loss of limiting nutrients, by the efficiency with which organisms use limiting resources, and by the proportion of limiting resources that the organisms living in an ecosystem are able to capture.

C. Functional Groups

Each species has a large number of morphological, physiological, and behavioral traits, many of which might influence the abundance of species and ecosystem functioning. One way to deal with such complexity has been to identify traits that seem more likely to influence ecosystem processes. Chapin *et al.* (1997) suggested that the species traits with the greatest effects on ecosystem functioning were those that (a) controlled the acquisition, use, and availability of limiting resources; (b) modified the feeding structure of food webs; and (c) affected the occurrence and magnitude of disturbances. Such traits can be used to classify organisms into different functional groups. For instance, species can be divided, first, into functional groups based on their position in a food web: photosynthetic plants, herbivores, predators, parasites, parasitoids, decomposes, and so on. Organisms within each of these groups can be further subdivided based on their acquisition and use of their limiting resources. For grassland plants, for instance, this might be based on the time, within the growing season, when each plant was maximally active (cool-season versus warm-season plants), and on its carbon (C-3 or C-4 photosynthetic pathway) and nitrogen physiology (high nitrogen use efficiency, low nitrogen use efficiency, ability to fix atmospheric nitrogen). Such considerations might lead, for instance, to the classification of grassland plants into six functional groups: C-3 grasses, C-4 grasses, C-3 forbs, C-4 forbs, legumes, and woody plants. The assumption inherent in making such a classification is that species within a class are highly similar, and those in different functional groups differ markedly from one another.

D. Diversity versus Composition

It has long been recognized that the functioning of an ecosystem depends on which species the ecosystem contains (i.e., on it species composition). Interest in species diversity as an alternative or additional explanation for ecosystem functioning means that it is necessary to define species diversity, especially functional diversity, in a way that distinguishes diversity from species composition. This requires a definition that is more restricted than that traditionally used. In particular, effects should be attributed to diversity only once there has been simultaneous control for effects of composition, and effects should be attributed to composition only once there has been simultaneous control for effects of diversity. To achieve this in an experimental, theoretical, or observational study, it is necessary (a) to hold composition constant via randomization (numerous communities with randomly-chosen compositions) while changing diversity, (b) to hold diversity constant while changing composition, (c) to simultaneously vary both in an appropriately randomized and replicated design, or (d) to control for each statistically, such as via multiple regression, which is most appropriate for observational studies.

III. EARLY WORK ON FUNCTIONAL DIVERSITY AND ECOSYSTEM PROCESSES

Effects of diversity on ecosystem processes were first recognized by Darwin in *The Origin of Species*. Darwin noted that it was well-known that increased plant diversity led to greater primary productivity in pastures. The British ecologist, Charles Elton, hypothesized in his 1958 book titled *The Ecology of Invasion by Animals and Plants* that diversity would impact many aspects of ecosystem functioning. In particular, he suggested that greater diversity would lead to greater ecosystem stability, an idea that was further developed by the leading ecologists of that era, including Robert MacArthur, Gene Odum, and Ramon Margalef. Elton also suggested that greater diversity would decrease the susceptibility of an ecosystem top invasion by other species and would decrease the incidence of outbreaks by diseases and pests.

Elton's diversity-stability hypothesis was called into question, though, by the mathematical theory of May (1972), which predicted that the linear stability of communities of competing species would, in general, decrease as the diversity of the communities increased. The general consensus reached after publication of May's book was that other factors were likely to be more important than diversity as determinants of ecosystem processes. This view led ecologists to focus more of their attention on other issues, with much of that effort dedicated to better understanding the mechanisms of species interactions and the effects of species composition on ecosystem processes.

Recent explorations of the potential effects of diversity on ecosystem processes were inspired, to a great extent, by the publication of *Biodiversity and Ecosystem Functioning* (Schulze and Mooney, 1993). In a chapter in that book, Vitousek and Hooper hypothesized that many ecosystem processes, like primary productivity, should increase as diversity increased, and they stressed that the most important component of diversity might be functional group diversity. Agricultural studies were reviewed in a chapter by Swift and Anderson, who noted that mixed crops, especially those containing a legume and a grass, were often more productive than either crop species growing alone, supporting the diversity-productivity hypothesis. A chapter by McNaughton reviewed and evaluated a large number of observational and small-scale experimental studies in which stability was greater for ecosystems containing more species and highlighted data supporting Darwin's diversity-productivity hypothesis. These and other contributions in this book set the stage for a burst of work that has included development of additional mathematical theories, field and laboratory experiments, and observational studies.

IV. THE EFFECTS OF FUNCTIONAL DIVERSITY

A. Functional Diversity, Productivity, and Nutrient Dynamics

1. Theory and Concepts

The potential effects of functional diversity on productivity have been described by two qualitatively different models, reviewed in Tilman (1999). The first is the sampling effect model, simultaneously proposed in 1997 by three different authors (L. Aarssen; M. Huston; and D. Tilman, C. Lehman, and K. Thomson). The sampling effect model hypothesizes that species differ in their competitive abilities, and that species that are better competitors are also more productive. Given these assumptions, communities that have greater diversity should, on average, be more productive because they are more likely to contain one or more species that are more productive.

A formal mathematical treatment of the sampling effect, provides some deeper insight into the way that functional diversity can impact ecosystem processes. For this treatment, let R^* be the level to which a limiting resource is reduced by a species when growing alone. As shown both theoretically and in numerous competition experiments (Grover 1997), the best competitor would be the species with the lowest R^*. The R^* value of the species can be used to rank them from good to poor competitive ability (i.e., from the lowest to the highest R^* value). Assume that the species composition of a community is determined by random draws (sampling) from the infinite pool of species with all possible R^* values between a minimum ($R^*_{\min}$) and maximum ($R^*_{\max}$). On average, the functional diversity of a community would depend on the number of species drawn, N, which is the initial diversity. The number of species in a community, N, is a good measure of functional diversity in this model because the range in the values of the relevant species trait (R^*) is higher in communities containing more species. These assumptions of sampling effect yield a simple equation that relates the long-term average biomass of a plant community, $B_{(N)}$, to its original plant species diversity, N:

$$B_{(N)} = aQ\left(S - \left(R^*_{\min} + \frac{R^*_{\max} - R^*_{\min}}{N+1}\right)\right) \qquad (1)$$

Here a is the rate of resource mineralization, Q the coefficient of resource conversion into biomass, and S is the rate of resource supply in the habitat.

The sampling effect model predicts that total community biomass, a measure of primary productivity, increases with plant diversity, as shown in Fig. 1a. The trend predicted is one in which added diversity leads to large increases in productivity when diversity is low, but has progressively smaller impacts when diversity is higher. This simple model demonstrates that the magnitude of the effect of functional diversity, as measured by N, on ecosystem functioning depends on the range of interspecific differences in the species pool—that is, on the term ($R^*_{\max} - R^*_{\min}$) in Equation 1. This gives basis to the intuitive concept that diversity effects ecosystem processes because ecosystems with greater diversity have a greater range in those species traits that influence functioning.

The sampling effect model also predicts that the average quantity of unconsumed resource should decrease as diversity increases (Fig. 1b). Indeed, in the sampling effect model, the increased biomass at higher diversity is caused solely by the more complete utilization of the limiting resource that occurs, on average, at higher diversity.

The model also illustrates the importance of species composition. Each point in the two graphs of Figure 1 represents the response of a community with a different randomly determined species composition. Thus, the variability among plots with the same diversity measures the impact of composition, and the variability among diversity levels represents the impact of diversity. Both diversity and composition are strong determinants of productivity and resource levels in the sampling effect model.

The other major type of models that have been proposed to relate productivity to diversity are niche differentiation models. In essence, such models assume that a habitat is spatially or temporally heterogeneous, that species differ in the traits that determine their response to this heterogeneity, and that each species is a superior competitor, and thus is more productive, for some subset of the heterogeneous habitat conditions. These assumptions can allow a large number of species to coexist and assure that ecosystem productivity increases, on average, as diversity increases. For instance, two factors, such as soil pH and temperature, might limit plant abundance. Each species could have some combination of these factors at which it performed best. Such niche

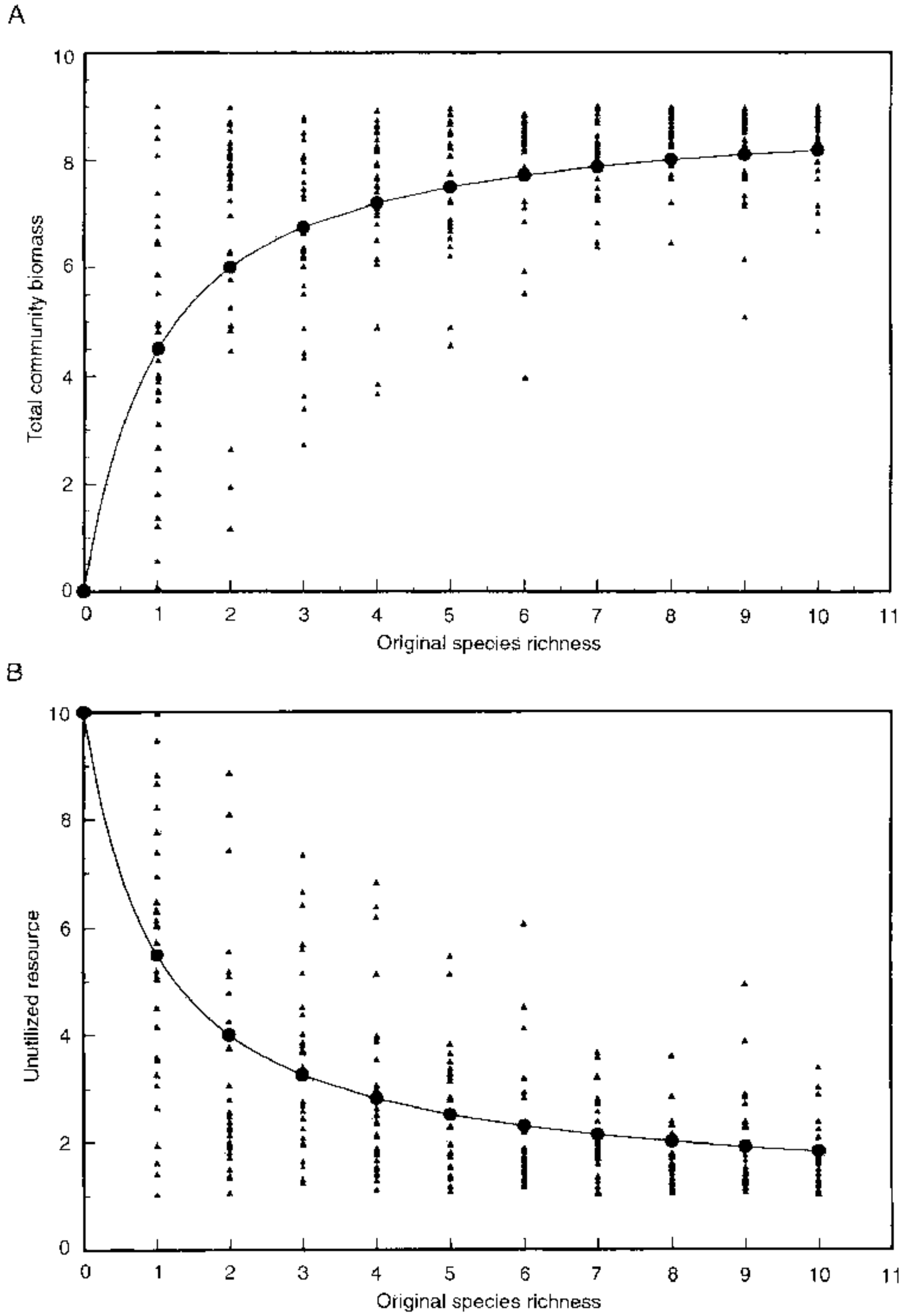


FIGURE 1 (A) The sampling effect model predicts that productivity should be greater at greater functional diversity, here measured by the number of species present. The variation within a given level of species richness is caused by different species compositions. (B) Productivity is higher in plots with greater functional diversity because of greater capture of the limiting resource. The concentration of unutilized resource is predicted to decline as diversity increases.

differentiation would mean that each species did best in a part of the habitat, but that no species could fully exploit the entire range of conditions.

The essence of such niche models can be captured by making the simple assumptions that each species has a circular area of radius r in which it can live and be a good competitor (Fig. 2a), that all species attain comparable abundances per unit habitat occupied, and that competition similarly reduces abundances of all overlapping species. If the values for one limiting factor range from 0 to $a \cdot r$ and the other from 0 to $b \cdot r$, where a and b measure habitat heterogeneity for factors 1 and 2, and if species are drawn at random from all those that could live at some point in the habitat, then total community biomass (i.e., the proportion of environmental conditions "covered" by one or more species) would be

$$B_{(N)} = 1 - \left(1 - \frac{\pi}{ab + 2(a + b) + \pi}\right)^{N} \qquad (2)$$

Here N is species diversity. $B_{(N)}$ is an increasing function of species diversity (Fig. 2b). The amount of unused habitat decreases as diversity increases, much as the concentration of unutilized resource was decreased for the sampling effect model. As for the sampling effect model, the variance within a given level of diversity is caused by differences in species composition, and differences between diversity levels is caused by diversity.

In addition, the niche model predicts that greater habitat heterogeneity (i.e., greater values of a and b) requires greater diversity in order to achieve a given level of productivity. In general, heterogeneity should increase with habitat size, leading to the prediction that greater biodiversity is required to attain a given level of productivity in larger habitats. For instance, for small, relatively homogeneous habitats ($a = b = 1$), only six species are needed to attain 95% of maximal productivity. However for spatially heterogeneous habitats ($a = b = 10$), a diversity of 135 plant species is needed to achieve this level.

A comparison of the sampling effect model with the niche differentiation model reveals a major difference in the expected pattern of the dependence of productivity on diversity. For the sampling effect model, there are no higher diversity plots that are more productive than the most productive monoculture. In contrast, for the niche model, there are two-species plots that are more productive than the most productive monoculture, three-species plots that are better than the best two-species plot, and so on. For ecosystems that meet the assumptions of the sampling effect model, which might occur for highly productive agricultural fields, there might be situations in which judicious choice of the right species and variety could lead to as great productivity from a monoculture as would be possible for a highly diverse mixture of species. In contrast, for habitats with spatial or temporal heterogeneity, which should occur for almost all natural ecosystems and for all but the most intensively managed ecosystems, niche differentiation models are more likely to hold. In such cases, increased diversity is expected to lead to greater productivity and to more complete use of limiting resources.

Although these models, and the models of Michel Loreau, have predicted that greater diversity can lead to greater ecosystem productivity, this need not always be the case. For instance, if the assumptions of the sampling effect model were modified to have progressively better competitors be progressively less productive, productivity would be a decreasing function of diversity. This suggests a more general principle: if species differ in their competitive abilities, and if higher competitive ability is correlated with some other traits, then these traits will, on average, be better represented in more diverse communities, thus biasing the functioning of these communities in the direction determined by these correlated traits.

2. Experimental Studies

Darwin suggested that it was common knowledge among farmers that a greater diversity of pasture plants would lead to a greater production of herbage in pastures. In his 1993 chapter, McNaughton cited this and presented more recent examples in which greater plant diversity led to greater productivity, as did Swift and Anderson. Indeed, earlier work reviewed in Harper's 1977 book showed that pairs of coexisting species often yield more than either species did when living by itself. As reviewed in the 1993 chapter by Vitousek and Hooper, some of the first evidence linking higher plant diversity to greater retention of soil nutrients came from a field experiment in Costa Rica by Ewel as collaborators. They found that communities planted to many tropical species generally retained more soil fertility than those planted to monocultures.

The first published direct test of the diversity-productivity hypothesis came from a greenhouse experiment by Naeem *et al.* (1995). By growing various randomly chosen combinations of 16 plant species 1, 2, 4, 8, or 16 at a time in a greenhouse, they found that community biomass was greater at higher plant diversity (Fig. 3a). This team performed another experiment in a series of growth chambers and also had results

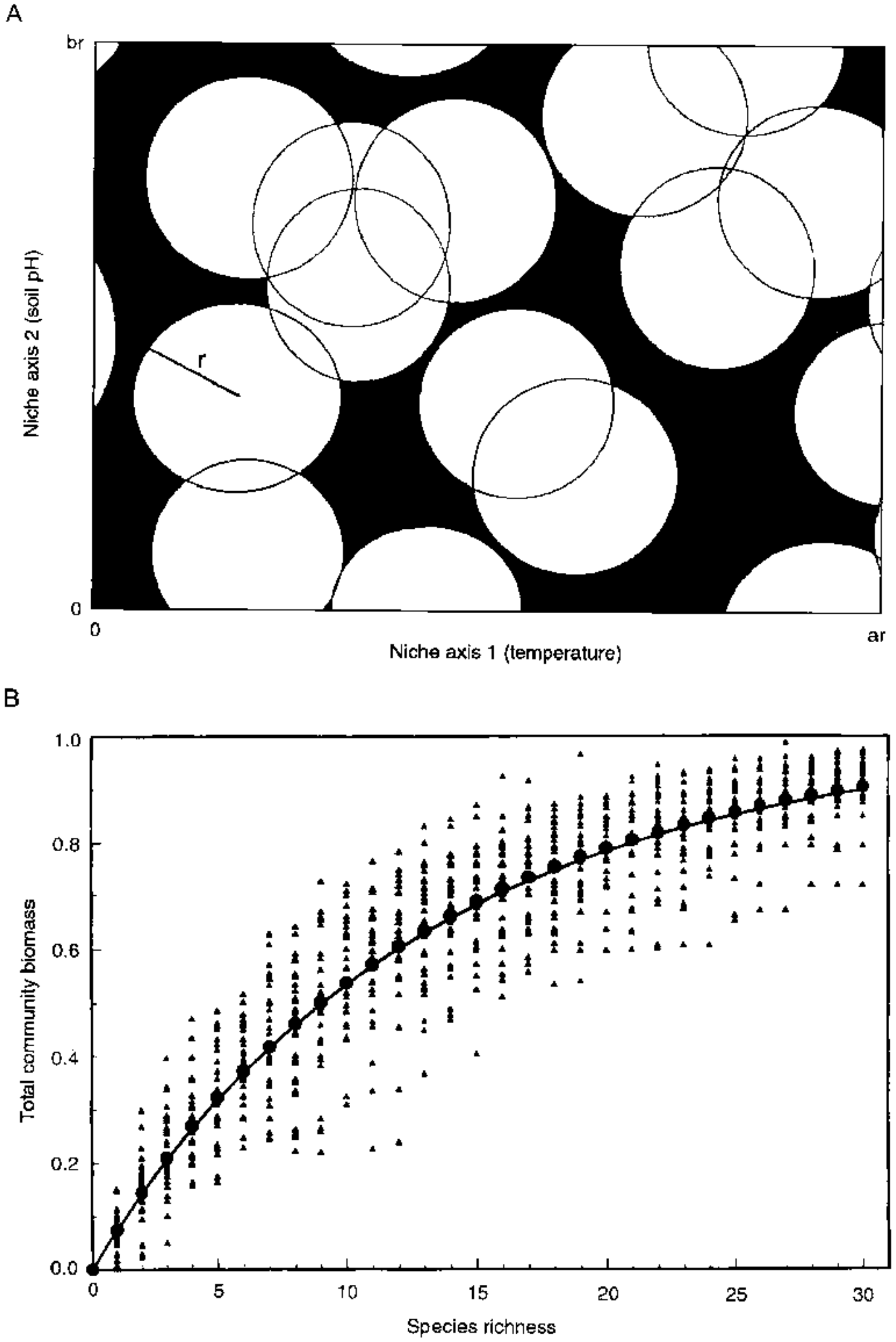


FIGURE 2 (A) A graphical illustration of a niche differentiation model. Here each circle represents the range of environmental conditions in which a given species can live, and the full rectangle shows the range of environmental conditions that occur in a given habitat. This model and similar niche differentiation models predict that productivity should be an increasing function of diversity. (B) The predicted effects of diversity on productivity for the model illustrated in part (A).

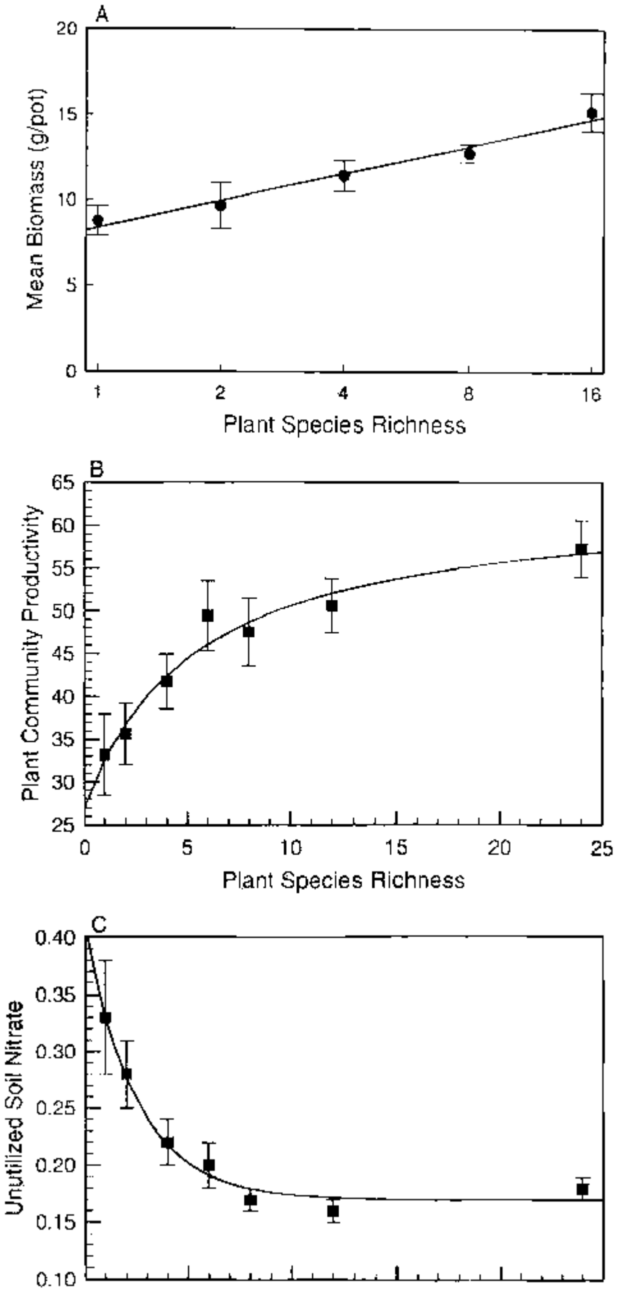


FIGURE 3 (A) The observed effect of plant diversity on the productivity of plant communities in the greenhouse experiment of Naeem and collaborators. (B) Effects of diversity on productivity for the Minnesota field experiment in which grassland diversity was experimentally controlled in 147 plots. (C) Effects of diversity on the concentration of unutilized soil nitrate for the Minnesota experiment.

suggesting that greater diversity leads to higher productivity (Naeem *et al.*, 1994). Next came results from a large-scale field experiment begun in Minnesota in 1993 (Fig. 4). Its 147 plots, each 3 m × 3 m, were planted to contain 1, 2, 4, 6, 8, 12, or 24 plant species randomly and independently chosen from a set of 24 prairie-grassland species (reviewed in Tilman, 1999). It found highly significant effects of plant diversity on both productivity (Fig. 3b) and on the soil concentration of the limiting resource, nitrate (Fig. 3c). By the fifth year of this experiment, its results supported niche differentiation models more than the sampling effect model as the major cause of the effects of diversity on the measured ecosystem processes. Indeed, the most productive plot in 1998 was a 24-species plot that had 65% greater total biomass than the most productive monoculture. A second experiment, adjacent to this Minnesota exper-

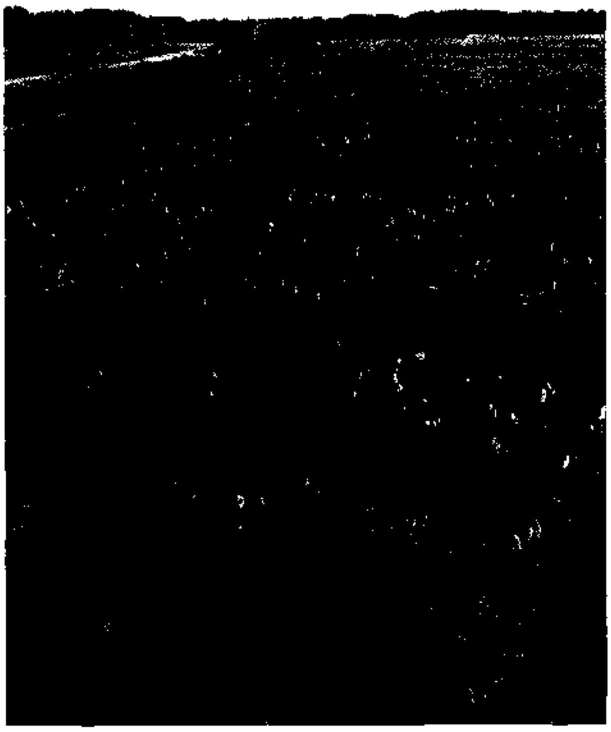


FIGURE 4 The smaller of the Minnesota biodiversity experiments, shown here, has demonstrated that plant diversity has a strong effect on ecosystem productivity and nutrient dynamics. The experiment has 147 plots, each being 3 m × 3 m (about 10 feet by 10 feet) in size. See also color insert, Volume 1.

iment (reviewed by Tilman, 1999), controlled for both species diversity and functional group diversity (Fig. 5). Its results were similar to those of the first experiment and showed highly significant effects of species diversity, functional group diversity, and functional group composition on primary productivity and nutrient dynamics. In both of the Minnesota grassland diversity experiments, the vast majority of species coexisted in all plots to which they had been added, further supporting niche differentiation models.

Knops *et al.* (2000) recorded the number of nonplanted species that invaded the Minnesota diversity experiment plots, and their biomass at the time when they were removed from the plots. They found that significantly fewer species invaded higher diversity plots and that the total biomass of invading species was lower in higher diversity plots. Further analyses suggested that the effect of diversity on invasions was caused by the lower levels of soil nitrate in higher diversity plots. This provides one simple mechanism whereby diversity may influence the extent to which an ecosystem is invaded by other species and suggests that levels of unconsumed limiting resources may, in general, be an important determinant of the success of an invading species.

For native, undisturbed grasslands close to the two Minnesota biodiversity experiments, plant abundances were greater and soil nitrate was lower in more diverse plots (see Tilman, 1999), which is consistent with the experimental results and with the predictions of theory. However, correlational patterns must be interpreted carefully because they could be confounded by other correlated variables. Michel Loreau used a model that linked environmental factors, biodiversity, and ecosystem functioning to explore this point. The model illustrated that correlational field data could be misinterpreted easily because of a confusion of cause-and-effect relationships. Just such issues cloud the interpretation of the possible effects of island diversity on ecosystem processes for a study of 50 Swedish islands. In an intriguing study that showed links between island size and the frequency of wildfire, David Wardle and collaborators found that a suite of ecosystem traits were correlated with both island size and plant diversity. However, it is unclear if diversity caused the observed differences in ecosystem processes or if both these processes and diversity were controlled by fire frequency.

Hooper and Vitousek (1998) performed a field experiment, planted in 1992, in which they controlled plant functional group diversity and composition using plants common to California grasslands. After a year of growth, they found that functional group composition had a much greater effect on plant community biomass than functional group diversity, but that the utilization of soil nutrients increased significantly as diversity increased.

FIGURE 5 The larger of the Minnesota biodiversity experiments uses about 270 of the 342 plots shown. It has shown strong effects of plant species richness, plant functional group richness, and plant functional group composition on ecosystem processes. Each plot is 13 m × 13 m (about 40 feet by 40 feet). See also color insert, Volume 1.

In a 4-month greenhouse experiment, Symstad *et al.* (1998) found that total plant biomass was significantly higher at higher diversity and that most of this effect was attributable to the presence of legumes. They also determined the effects of the deletion of individual species on total biomass and found that the strength and direction of these effects depended on which species were present and which was deleted.

In an experiment that was replicated at eight different sites across Europe, ranging from Scotland and Ireland to Portugal and Greece, Hector *et al.* (1999) found that greater plant diversity led to greater primary productivity. An important finding of this unique experiment was that the quantitative effect of diversity on primary productivity was the same across all eight sites. In combination with the other field and laboratory experiments, the European experiment suggests that there is a general, repeatable effect of grassland diversity on primary productivity.

In total, these studies show that plant productivity is greater at greater diversity and that this also corresponds with greater utilization of limiting soil resources. In general, short-term experiments showed weaker effects of diversity on productivity and soil nutrients than longer-term experiments. This is expected because diversity should impact ecosystem processes via changes in plant abundances mediated by competition, and such interactions can require several years to occur. Further work is needed on other trophic levels and in other communities to determine the extent to which the patterns observed to date apply to other trophic levels (e.g., herbivores, predators) or to other communities (e.g., marine fisheries, forest ecosystems, coral reefs).

B. Functional Diversity and Stability

1. Theory and Concepts

A large number of authors, including Charles Elton, Robert May, Stuart Pimm, and Sam McNaughton have contributed considerable insights into the effects of diversity on stability. May (1972), for instance, showed that the abundances of individual species become progressively less stable as the diversity of the community in which they live increases. Several recent papers have explored the effects of diversity on the stability of communities of competing species (Doak *et al.*, 1998; Ives *et al.*, 1999; Tilman, 1999). The first two of these papers showed that the temporal variability of an ecosystem process, such as ecosystem productivity, is expected to be lower when the ecosystems contain more species. This can occur for the same reason that a portfolio composed of many different types of stock tends to be more stable than one containing stock of a single company. An additional factor that can cause ecosystem functioning to be more stable for more diverse ecosystems is competition. When some disturbance harms one species, the species with which it interacts experience less competition. This allows these competitors to increase in abundance. Their greater abundance partially compensates for the decreased abundance of the first species, thus stabilizing the functioning of the ecosystem. Ives *et al.* (1999) showed that increased diversity only led to increased stability when the species differed in their responses to habitat fluctuations and disturbances. Because such differences are a direct measure of functional diversity, the work of Ives *et al.* (1999) showed that increases in functional diversity lead to greater stability. For a thorough treatment of theory relating diversity and stability, see "Stability, Concept of."

2. Experimental and Observational Studies

The evidence that led Elton to propose the diversity-stability hypothesis was anecdotal. In his 1993 chapter, and in earlier papers, McNaughton defended the diversity-stability hypothesis by citing several observations and experiments in which greater diversity was associated with greater stability. A variety of other studies, summarized in Tilman (1999), also have found effects of diversity on stability. For instance, a study by Frank and McNaughton of eight grassland sites within Yellowstone National Park found that those with greater plant species diversity had smaller shifts in plant community compositions during a severe drought. Two British ecologists, Taylor and Woiwod, performed a long-term project in which they monitored the abundances of hundreds of insect species at a large number of sites. The data they collected provide evidence that supports the hypothesis that more diverse insect communities should be more stable. The greater stability is expected because of the statistical averaging (or portfolio) effect pointed out by Doak *et al.* (1998). Specifically, because the temporal variances in the abundances of individual species in this community scales as their abundance to a power of about 1.6, the portfolio effect should cause more diverse insect communities to have lower temporal variability.

Several authors have found that greater oak tree diversity stabilizes the population density of an animal, the acorn woodpecker, that feeds on the seeds of the trees (see Koenig and Haydock, 1999). Acorn woodpeckers are highly dependent on acorns as a source of food, but oaks produce acorns as a mast seed crop. Masting means that there is great year-to-year variability in the rate of acorn production. There is a striking decrease in the year-to-year variability of acorn woodpecker abundances for woodpeckers living in habitats containing a greater diversity of oaks. Thus, greater oak diversity led to more stable acorn woodpecker populations. Moreover, acorn woodpecker densities were much lower for areas with a single oak species than for those with several.

A long-term experiment in Minnesota provides additional evidence suggesting that greater plant diversity leads to greater stability (reviewed in Tilman, 1999). In a series of 207 plots annually monitored from 1982 to 1999, total plant community biomass was found to be more stable in plots containing more species. Both in response to a major disturbance, a severe drought (Fig. 6), and in response to normal year-to-year variation in climate (Tilman, 1999), plots with greater diversity had lower year-to-year variability in their total plant biomass. In particular, the severe drought caused plant biomass to fall to half of its predrought level in plots with about 15 or more species, but caused it to fall to 1/8 to 1/12 of its predrought levels in plots containing one or two plant species (Fig. 6). Similarly, year-to-year variation in total biomass fluctuated about twice as much in low diversity as in high diversity plots (Tilman, 1999). Although total community biomass was more stable at higher diversity, analyses of the stability of individual species showed that these declined slightly but detectably, at higher diversity. Thus, diversity stabilized total community biomass at the same time that it destabilized the abundances of individual plant species. Plant diversity and composition were confounded in this experiment because both changed in response to nitrogen addition. Multiple regression, used to control for this confounding, found highly significant effects of diversity on stability for both cases. These analyses also showed that species composition and functional group composition also had significant effects on stability.

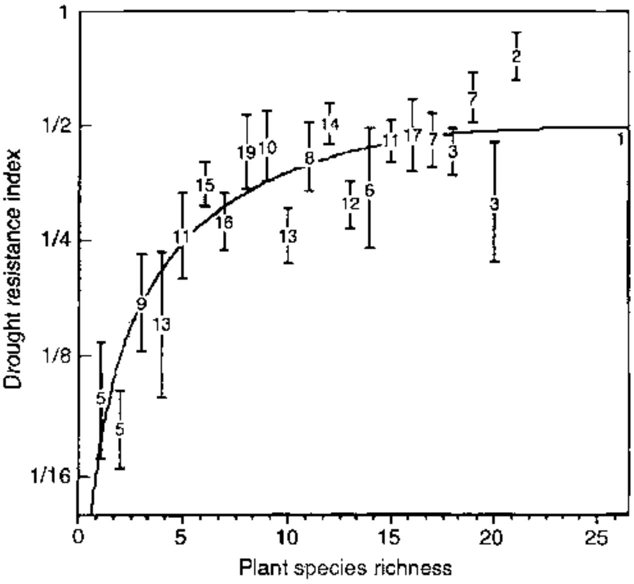


FIGURE 6 The resistance of Minnesota grassland ecosystems to drought was highly dependent on their plant biodiversity. Ecosystems containing a large number of plant species had their productivity fall to about half of its predrought levels during a severe drought, but those containing only one or two plant species had it fall to about 1/8 to 1/12 of the predrought level.

McGrady-Steed, Harris, and Morin (1997) found, in a laboratory study of the effects of diversity in microbial communities, that the temporal variability was significantly smaller at higher diversity. Indeed, a four-fold increase in diversity led to about a three-fold decrease in the temporal variability of whole-community net respiration, a measure of ecosystem activity. The rate of microbial decomposition of particulate organic matter also increased with diversity in this study. Finally, they found that greater diversity led to lower susceptibility to invasion by another species, but that invader success was highly dependent on community composition. Naeem and Li (1997) similarly found that greater diversity led to greater reliability, which was measured as the lower variability in total community biomass among communities of identical diversity. This effect was also apparent in the greenhouse experiment that Naeem and collaborators had performed earlier.

In total, these studies provide strong evidence that communities with greater diversity are more stable and suggest that individual species in such communities may be less stable. Theory, experiment, and observation are in general agreement, but this topic merits additional exploration.

V. CONCLUSIONS

The research performed to date illustrates that a variety of different ecosystem processes are impacted by the number and kinds of species living in the ecosystem.

This work illustrates that species differ in traits that influence ecosystem functioning and suggests that ecosystem processes depend on the range in those traits represented in the ecosystem. However, there are, as yet, no clear demonstrations of the specific traits that are relevant to particular ecosystem processes and no simple ways to directly measure functional diversity. Rather, correlates of functional diversity, such as species richness or functional group richness, remain the best, albeit indirect, way to measure functional diversity.

See Also the Following Articles

C_4 PLANTS • ECOSYSTEM FUNCTION, PRINCIPLES OF • FUNCTIONAL GROUPS • HABITAT AND NICHE, CONCEPT OF • STABILITY, CONCEPT OF

Bibliography

Chapin, F. S., III, Walker, B. H., Hobbs, R. J., Hooper, D. U., Lawton, J. H., Sala, O. E., and Tilman, D. (1997). Biotic control over the functioning of ecosystems. *Science* 277, 500–504.

Doak, D. F., Bigger, D., Harding, E. K., Marvier, M. A., O'Malley, R. E., and Thomson, D. (1998). The statistical inevitability of stability-diversity relationships in community ecology. *Am. Nat.* 151, 264–276.

Grover, J. P. (1997). *Resource Competition*. Chapman & Hall, London.

Hector, A., Schmid, B., Beierkuhnlein, C., Caldeira, M. C., Diemer, M., Dimitrakopoulos, P. G., Finn, J., Freitas, H., Giller, P. S., Good, J., Harris, R., Högberg, P., Huss-Danell, K., Joshi, J., Jumpponen, A., Körner, C., Leadley, P. W., Loreau, M., Minns, A., Mulder, C. P. H., O'Donovan, G., Otway, S. J., Pereira, J. S., Prinz, A., Read, D. J., Scherer-Lorenzen, M., Schulze, E.-D., Siamantziouras, A.-S. D., Spehn, E. M., Terry, A. C., Troumbis, A. Y., Woodward, F. I., Yachi, S., and Lawton, J. H., (1999). Plant diversity and productivity experiments in European grasslands. *Science* **286**, 1123–1127.

Hooper, D. U., and Vitousek, P. M. (1998). Effects of plant composition and diversity on nutrient cycling. *Ecological Monographs* 68(1), 121–149.

Ives, A. R., Gross, K., and Klug, J. L. (1999). Stability and variability in competitive communities. *Science* **286**, 542–544.

Knops, J. M. H., and Tilman, D. (2000). Dynamics of soil nitrogen and carbon accumulation for 61 years after agricultural abandonment. *Ecology* 81, 88–98.

Koenig, W. D., and Haydock, J. (1999). Oaks, acorns, and the geographical ecology of acorn woodpeckers. *Journal of Biogeography* 26, 159–165.

May, R. M. (1972). Will a large complex system be stable? *Nature* 238, 413–414.

McGrady-Steed, J., Harris, P. M., and Morin, P. J. (1997). Biodiversity regulates ecosystem predictability. *Nature* 390, 162–165.

Naeem, S., and Li, S. (1997). Biodiversity enhances ecosystem reliability. *Nature* 390, 507–509.

Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H., and Woodfin, R. M. (1994). Declining biodiversity can alter the performance of ecosystems. *Nature* 368, 734–737.

Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H., and Woodfin, R. M. (1995). Empirical evidence that declining species diversity may alter the performance of terrestrial ecosystems. *Philos. Trans. Royal Soc. London B.* 347, 249–262.

Schulze, E. D., and Mooney, H. A. (1993). *Biodiversity and Ecosystem Function*. Springer Verlag, Berlin.

Symstad, A. J., Tilman, D., Willson, J., and Knops, J. M. H (1998). Species loss and ecosystem functioning: Effects of species identity and community composition. *Oikos* **81**, 389–397.

Tilman, D. (1999). The ecological consequences of changes in biodiversity: A search for general principles. *Ecology* 80, 1455–1474.

FUNCTIONAL GROUPS

Robert S. Steneck
University of Maine

GLOSSARY

convergent evolution Distant or unrelated organisms evolve the same anatomical, morphological body plan characteristics, or ecological function.

ecosystem The combined plant and animal communities plus their physical environment.

ecosystem function The energy flow, productivity, element cycling, and resilience of ecosystem structure; synonymous with "behavior of ecological systems" and "ecological processes."

ecosystem structure The organisms, their communities, biodiversity, and habitats that comprise an ecosystem.

functional groups Polyphyletic suites of species that share ecological characteristics and play equivalent roles in natural communities and ecosystems. Commonly, organisms with convergent anatomical, morphological, physiological, behavioral, biochemical, or trophic characteristics are grouped together.

guilds Organisms that use similar resources in similar ways. Depending on the application, guilds can be synonymous with functional groups.

EARTH'S BIODIVERSITY IS DISTRIBUTED among a surprisingly few functionally different organisms. These "functional groups" are suites of species that play equivalent roles in natural communities and ecosystems. They result from convergent evolution which is channeled by phyletic constraints that limit the variations possible on a given body plan relative to the biomechanical limits imposed on organisms by their environment. The diversity and ecological function of these groups are independent of regional species diversity. Large-scale changes in functional groups can alter the structure and functioning of ecosystems. Such alterations have occurred over evolutionary time and are occurring today in ecosystems under strong human influences. In highly diverse ecosystems there are redundant species within functional groups. These redundancies provide a buffer against ecosystem collapse should individual species within the group become rare or extinct. Describing communities and ecosystems at the functional group level conveys information that transcends taxonomic and evolutionary boundaries. Ultimately, functional groups may provide a low-resolution tool for accurately predicting ecosystem change.

I. FUNCTIONAL GROUPINGS AND THEIR UTILITY

More than 40 years ago, ecologist and evolutionary biologist G. E. Hutchinson wondered why there were so many kinds of species. This question engaged many of the most powerful minds in ecology and the natural sciences. It launched decades of research, brought us the concept of ecological niche and considerable ecological and evolutionary theory, and is the focus of many articles in this encyclopedia. Much less has been written about an equally compelling question: Why are there so few kinds of morphologies or physiologies, and what might the ecological consequences of such limits be? It also raises new questions, such as the following: Are there limits to the possible number of variations on any given body plan? Are there similar limits to the functional attributes of species and suites of species? Can functional attributes of species be abstracted such that their ecological roles can be inferred? Can several species fulfill the same functional roles? In other words, are species replaceable such that there are ecologically equivalent or redundant species? The answers to these questions have profound implications to how we view the natural world and possibly how we can conserve it. For example, does the extinction or extirpation of any given species degrade natural ecosystems in tangible ways?

Biodiversity includes the diversity of organic life on Earth and how it functions. How organisms "make their living," how they interact and contribute to the ecosystem, relates to the functional properties of biodiversity. Functional groupings are polyphyletic suites of species that share ecological characteristics. Commonly, organisms with convergent anatomical, morphological, physiological, behavioral, biochemical, or trophic characteristics are grouped together. Functional properties may correspond to different ecological characteristics depending on the question being asked. They can include the habitats in which they live, the biomass or structure they impart to a community, the rates at which they grow (their productivity), the rates at which they lose biomass (i.e., disturbance), the organisms they consume, the organisms that consume them, and their numerical or mass-specific importance to their communities or ecosystems. A functional group analysis can be applied more broadly in space for making biogeographic comparisons and in time for reconstructing paleocommunities than is possible at the level of species or of related taxa.

I explore functional groupings first from the perspective of the evolution of functional similarities, and then I discuss what functional groupings might mean to the structure and function of communities and ecosystems. I draw heavily, but not exclusively, on examples from the marine realm, in which there is the greatest phyletic diversity, longest evolutionary history, and clearest fossil record of how organisms and assemblages have evolved. Other extensive works on the topic of functional groups, such as that by Schluze and Mooney (1994a), are exclusively terrestrial in their approach.

Functional distinctions can be based on any characteristic shared among organisms, such as morphology, physiology, or behavior. These are not independent features of organisms. An organism's morphology can influence its behavior and its physiology. Functional groupings would be unnecessary if organisms' taxonomic relationships corresponded well with their functional roles. However, functional characteristics are often decoupled from even closely related species, and thus the relationship between taxonomic relatedness and functional groupings warrants consideration.

II. DECOUPLING RELATEDNESS FROM FUNCTIONAL SIMILARITIES

Species that are closely related evolutionarily are not necessarily morphologically or functionally similar. Conversely, just because organisms are distantly related does not mean they are morphologically or ecologically different. There are examples of unrelated, or distantly related, organisms having similar morphologies, anatomies, behaviors, physiologies, and functional roles in natural communities. Such convergent functional properties are well-known and operate at many levels. Obviously, bats are flying mammals that exploit some of the same habitats and foods as some birds and moths. Squid are swimming mollusks that exploit some of the same habitats and foods as fish. Although bats are not perfect substitutes for birds, nor are squid identical to fish, their phyletic relationships tell us little about the ecological roles or functional characteristics they posses.

Although most closely related organisms share traits with the clade from which they evolved, and thus will tend to be functionally similar, this is not always the case. For example, hermit crabs are relatively small, anomuran decapods with an exposed, asymmetric abdomen that requires protection from unoccupied snail shells. They differ morphologically from true crabs, the brachyurans that have abdomens protected by a shield-like cephalothorax or carapace that also covers their

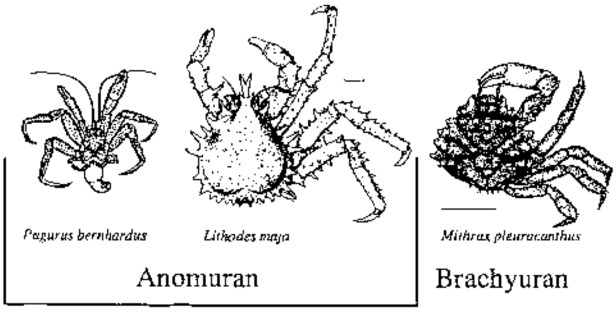


FIGURE 1 Phylogenetically related anomuran hermit crab (*Pagurus*) and king crab (*Lithodes*) are morphologically dissimilar. The distantly related brachyuran crab (*Mithrax*) is morphologically similar to the king crab (after Cunningham *et al.*, 1992).

walking legs (Fig. 1). True crabs can attain large size. However, hermit crabs are very closely related to the morphologically distinct anomuran king crab (Fig. 1). King crabs convergently evolved many of the characteristics of true brachyuran crabs; they are large and their abdomen is protected by a broad shield-like cephalothorax. Therefore, phyletically related king and hermit crabs are morphologically different, whereas phyletically distant king and true crabs are morphologically similar.

Morphological similarities can translate to functional similarities. In shallow coastal habitats of the Gulf of Maine, there are four species of large, clawed, decapods that live together and feed on the edible blue mussel, *Mytilus edulis* (Fig. 2). The two most closely related species are *Cancer irroratus* and *Cancer borealis*. Both are brachyuran crabs in the family Cancridae. The third brachyuran crab, *Carcinus maenas*, is in the Portunidae family. However, the two distantly related crabs, *Carcinus maenas* and *C. irroratus*, share functional characteristics of small body and chelae ("claws") size while being behaviorally flexible, quick, and dexterous. They characteristically chip the perimeter of mussel shells at many locations to gain access to the food. Conversely, *C. borealis* and the lobster *Homarus americanus* share functional characteristics of large body and chelae size while being behaviorally constrained, slow, and strong (Fig. 2). These two decapods characteristically crush mussel shells in the middle of the shell by applying brute force to the center of the shell. Therefore, the two most closely related crabs are functionally dissimilar and the two sets of most distantly related crabs (*Carcinus maenas* and *C. irroratus*) and other decapod groups (*C. borealis* and *H. americanus*) are both functionally similar.

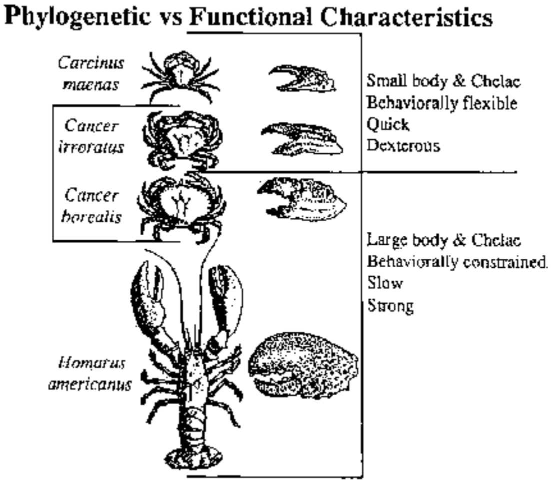


FIGURE 2 Phylogenetically related *Cancer* species are functionally different. The phylogenetically distant *Cancer borealis* and *Homarus americanus* are functionally similar (after Moody and Steneck, 1993).

Even within species, large functional differences exist. Amphibians such as frogs begin life as tadpoles with diets and habitats different from those of adults. Ferns and algae have ecologically distinct haploid and diploid phases. The heteromorphic range in some marine algae is striking. For example, the brown (Phaeophyta) algal kelp *Macrocystis pyrifera* can exceed 40 m in length and forms forests in its sporophyte (2*n*) phase but is a microscopic filament in its gametophyte (1*n*) phase. Some red algae (Rhodophyta) alternate between encrusting and erect phases; some green algae (Chlorophyta) alternate between endolithic microscopic and filamentous phases. Some habitats only support one phase. Each of these different ontogenetic and ploidy phases has strikingly different ecological properties, and differences within species can be greater than some differences among species. Therefore, knowing the species does not always provide insight into how the organism functions ecologically.

There are distinct limits to variety in morphology or important ecological function that are independent of biodiversity. Some characteristics, such as body size, are biomechanically limited; therefore, depending on the biomaterials available to an organism (i.e., cellulose, chitin, calcium carbonate, or bone), their morphology and the environment in which they live may limit their ecological functions. These relationships are well described in works such as Thompson's (1966) *On Growth and Form*, Peter's (1983) *The Ecological Implications of Body Size*, or Niklas's (1994) *Plant Allometry*. Arguably, size and shape are prime determinants of many functional groupings of both plants and animals. Predators generally scale with, or are larger than, their prey. Thus, organisms constrained to small size are likely to be more vulnerable to predators and rarely if ever are "apex predators" in an ecosystem or community. Insects are a good example. They also illustrate the disconnect between species diversity and morphological variety. Insects account for most of Earth's biodiversity, but they are all relatively small. In contrast, most vertebrates are much larger (especially mammals and dinosaurs) and their biomass distribution does not overlap that of insects. Despite their long evolutionary history, there are no cow-sized insects, nor are there insect-sized cows. Body size constraints among insects occur because, among other reasons, their respiration is relatively inefficient and limited by their open circulatory system and trachea. Other factors, such as the composition of their exoskeleton, further constrain morphological evolution and the habitats into which they can and have radiated. Although insects have radiated impressively into terrestrial systems, they are virtually nonexistent in the marine realm. Thus, three-fourths of all species are insects, but three-fourths of the habitable space is devoid of them. There are morphological and functional limits to diversity.

III. FUNCTIONAL CHANGES IN ASSEMBLAGES OVER EVOLUTIONARY TIME

Ultimately, the environment controls the distribution, abundance, and body plans that live in it. Evolution at several levels generates diversity of form and function, but the environment filters it. Constraints to diversification are many and relate to the environment as it changes and how organisms deal with it.

The history of life on Earth is punctuated with significant functionally novel additions to biotic assemblages. An excellent example of a functional group effect on global-scale ecosystem function is the cyanobacteria that began oxygenating the atmosphere more than 3 billion years ago. Biologically available oxygen began increasing approximately 1.4 billion years ago and is coincident with (and possibly necessary for) the evolution of the first metazoans. Still higher levels of oxygen may have been necessary for the Cambrian "explosion" of small shelly fauna. The sudden polyphyletic and global appearance of hard skeletal material (chitin and calcium carbonate) allowed for the evolution of large body sizes, efficient mobility, teeth, and defenses against predators. The structure and functioning of life on Earth changed at this time. Organisms functionally similar to those alive today first appeared. Since then, during the past 600 million years, there have been several biotic revolutions.

Decoupling phyletic evolution from functional group evolution is important. All but one of the major phyla may have evolved at or near the Precambrian–Cambrian boundary in as few as 25 million years. Alternatively, some argue that phyla evolution may have occurred much earlier (1 billion years ago) as minute and embryonic or larval-like organisms and thus not recognizable by today's criteria. If this latter theory is correct, it would be a clear case of phyletic evolution being disconnected from the functional/morphological (phenetic) evolution that occurred 400 million years later. Despite the Cambrian (or earlier) origin of most major phyla, functional evolution and the filling of unexploited habitats ("ecospace" *sensu* Bambach, 1985) took considerably more time. Throughout the Paleozoic, morphological evolution and the filling of unoccupied habitats occurred episodically and polyphyletically, beginning with small surface deposit feeders and low-profile suspension feeders, progressing to greater depths into sediment and new epifaunal heights. The progressive utilization of three-dimensional space called "tiering" is thought to have contributed substantially to the global diversification in the early Paleozoic (Fig. 3). The diversifying continued into the pelagic realm and into the sediment habitats (infaunalization) and recently included the radiation of large-bodied predators.

This pattern of diversification is composed of three relatively distinct faunas, each characterized by its own characteristic dominant taxa (Fig. 3A), initial rates of diversification, and "equilibrium" diversity. Each fauna was functionally different. Surface deposit feeders such as trilobites and monoplacophorans dominated the Cambrian fauna, whereas the Paleozoic fauna radiated into and relied heavily on the pelagic realm. For example, the taller suspension-feeding crinoids of the Paleozoic fauna joined near-benthos suspension-feeding brachiopods, anthozoans, and stenolomate bryozoans and pelagic groups such as cephalopods and graptolites as dominants of this fauna. There were no known macroalgal grazers in the Paleozoic, and large stands of macroalgae were probably common. Sand habitats were virtually devoid of infauna. In contrast, the modern fauna included the rise to dominance of large invertebrate and vertebrate predators and herbivores. These included predatory neogastropod snails, malacostracan crab and lobsters, and at least four groups of swimming vertebrate predators among the bony fishes (Osteichthyes), dinosaurs (Reptilia), sharks (Chondrichthyes), and cetaceans (Mammalia, i.e., dolphins and whales). Most modern groups of fish (e.g., most modern reef

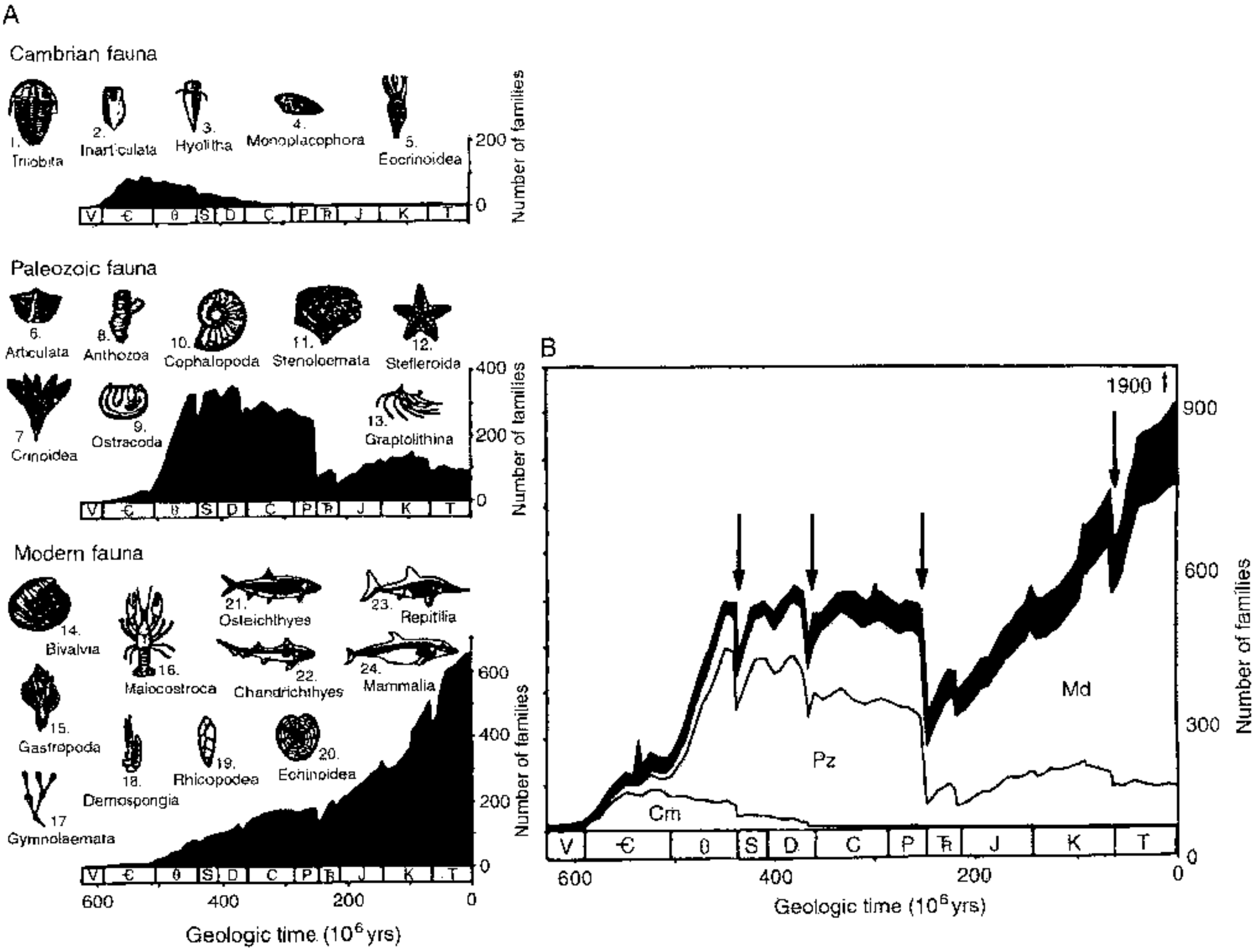


FIGURE 3 Faunal changes during the past 600 million years. (A) Three faunas with representative dominant animals. (B) Cumulative pattern of diversification of the three faunas. Arrows identify mass extinction events (after Sepkoski, 1984). Each fauna represents dominance by a functionally distinct group. The Cambrian fauna were surface deposit feeders, the Paleozoic fauna were dominated by suspension feeders, and the modern fauna were dominated by large predators and protected prey. Note that the Paleozoic fauna was particularly susceptible to mass extinctions resulting from disruptions of the water column.

fish) evolved at this time. Explanations for why large-scale changes occurred at the family level are complex. Organisms such as small deposit feeders may have suffered higher extinction rates as more mobile and more pelagic organisms with larger gene pools replaced them. A global crisis in the water column contributed to the most severe extinction event at the end of the Permian. Therefore, suspension feeders, pelagic organisms, and organisms with pelagic larvae suffered the highest extinction rates. This extinction caused a sharp decline in several of the dominant families of the Paleozoic fauna and marked the transition to the dominance of the modern fauna (Fig. 3B). Prior to the dominance of the modern flora there were virtually no shell-crushing carnivores. Several groups of shell crushers, such as true brachyuran crabs, evolved in the Cretaceous and became important predators. Deep-grazing herbivores also evolved among the modern fauna. They created a uniquely intense level of grazing pressure which may have denuded many shallow marine habitats of macroalgae.

A. The Sudden Origin of Functionally Distinct Organisms and Then Functional Stasis

Functional evolution in the marine realm is striking. The transitions from surface-dwelling deposit feeders to a range of suspension feeders, to organisms using vertical space, and to big predators and herbivores represent significant functional changes over evolutionary time. There were also terrestrial revolutions in organism function. Angiosperms displaced functionally similar conifers and evolved functionally unique weeds. Pollinating insects and angiosperms coevolved and diversified. Obviously, the history of life is punctuated with the sudden appearance of functionally unique organisms that often change the biotic world.

Often, functionally unique groups change little over evolutionary time. Among marine invertebrate herbivores, there are only a few capable of biting into limestone or calcium carbonate. These herbivores leave a grazing trace fossil that allows for functional interpretations to be made over evolutionary time. Three groups known to excavate calcium carbonate are regular echinoids (sea urchins), true limpets (Patellogastropoda), and chitons (Polyplacophora). Trace fossils from the Cretaceous (approximately 70 million years ago) reveal that each of these groups grazes distinctively with the same graze marks and to about the same depths into calcium carbonate as their modern counterparts (Fig. 4). Although considerable variation exists within groups of herbivores, the differences among groups are much greater. Echinoid sea urchins have a characteristic intensity of grazing that is much deeper than that of chitons and limpets which are very distantly related but have similar shallow bites into calcium carbonate. In all cases, the first trace fossils known show these invertebrates biting the surface of calcium carbonate with the pattern and the depth into the substrate as they produce today. There appears to be little or no functional change in how they graze after the group first evolved. There are other examples. Predatory gastropods that drill their shelled prey have not changed their mode of attack since they first evolved in the Triassic more than 200 million years ago. Many limb and running characteristics between terrestrial vertebrate predator–prey interactions have also been shown to be remarkably stable over evolutionary time. This suggests that among these consumers and within established clades, there is functional stasis over evolutionary time.

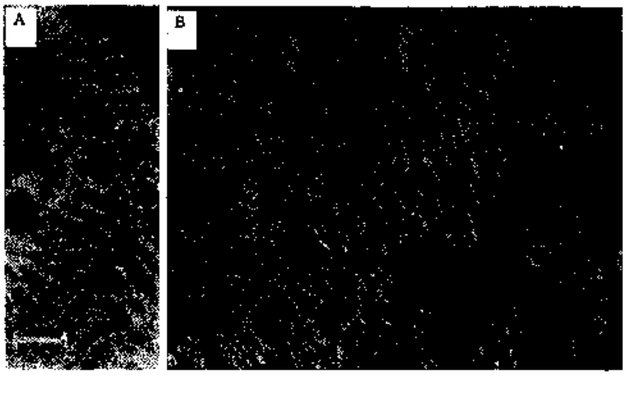


FIGURE 4 Comparison of modern and ancient limpet graze marks on calcium carbonate substrates. (A) Modern limpet, *Tectura testudinalis*, grazing calcareous algae (scale bar = 0.5 mm). (B) Cretaceous (more than 65 million years ago) limpet, *Radulichnus* sp. (scale is the same as in A) [reproduced with permission from (A) Steneck (1982) and (B) Crimes and Harper (1977)].

IV. FUNCTIONAL CONVERGENCES

Convergent evolution results in distantly related organisms converging on the same body plan or ecological function. Charles Darwin noted in his *Origin of Species* (1859) that

> for animals belonging to two most distant lines may have become adapted to similar conditions, and thus have assumed a close external resemblance; but such resemblances will not reveal—will rather tend to conceal—their blood relationship (page 463).

Relatively few convergently evolved anatomical, morphological, or physiological characteristics are the cornerstones of functional groupings.

A. Groupings among Mobile Organisms

Convergent evolution is well-known and documented in the terrestrial realm. Marsupial and placental mammals have converged to similar morphologies and ecological function (Fig. 5). The limited variations on the mammalian body plan are evident in the wolf and cat-like carnivores, the arboreal gliders, fossorial herbivores, anteaters, and subterranean insectivores that evolved independently in Australia for the marsupials and on the other continents for the placentals.

Convergent functions can be found among dissimilar-looking organisms. Distantly related marine molluskan herbivores, chitons and limpets, provide an excellent example. Chitons (class Polyplacophora) and snails (class Gastropoda) evolved in the Cambrian at the beginning of the molluskan diversification. The true limpets, Patellogastropoda, evolved from the Archaeogastropoda, much later in the Triassic. However, the radula (the teeth) of these two groups are functionally similar (Fig. 6). Both groups have relatively few teeth that contact the substrate, and those that do are hardened

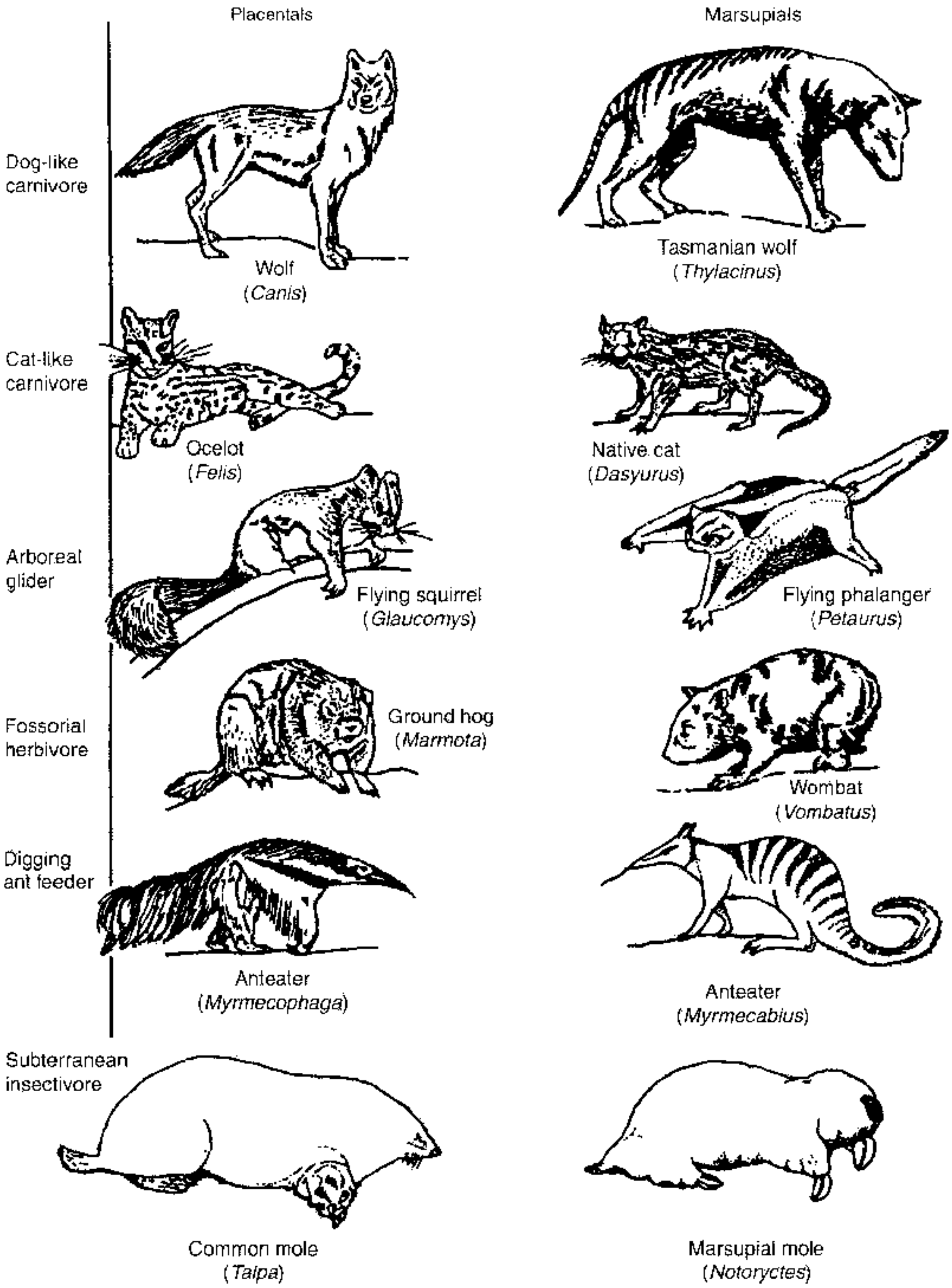


FIGURE 5 Convergent evolution among placental (left) and marsupial (right) mammals (reproduced with permission from Begon *et al.*, 1986).

by mineralization of iron or silica compounds (note the black teeth in Fig. 6). Both groups have strong buccal musculature for applying downward forces. Within the molluskan body plan, only chitons and limpets have such a large foot area-to-mass relationship and an excavating-type radula. These morphological and anatomical characteristics, along with their size and mobility, allow species of these groups to specialize on large and

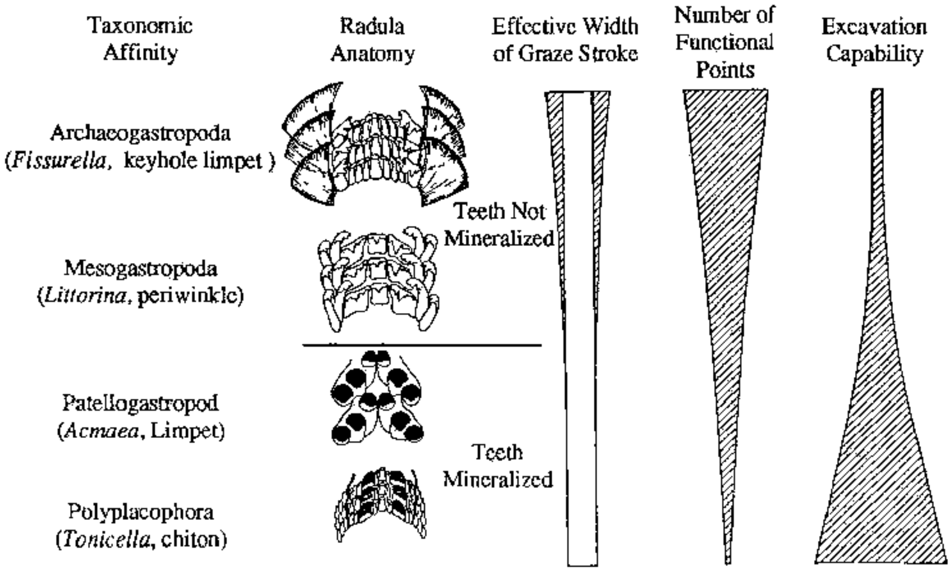


FIGURE 6 The teeth (i.e., radulae) of herbivorous mollusk functional groups. Black teeth are heavily mineralized, making them harder than the limestone they graze (after Steneck and Watling, Feeding capabilities and limitation of herbivorous molluscs. *Mar. Biol.* 68, 299–319, © 1982 Springer-Verlag, with permission).

expansive macrophytes such as sea grasses, kelp, and encrusting coralline algae. Although species diversity is much higher in other groups of mollusks such as nonlimpet gastropods, it is the functional characteristics of these two groups that make them capable of consuming and even trophically specializing on the tough or limestone-imbedded cells of kelp or coralline algae, respectively.

Convergent functions are numerous among terrestrial organisms. For example, a diversity of flowering plants and their pollinators possess similar morphological and anatomical characteristics despite significant phyletic separation among the plants and the pollinators. The geometry of the flowers, such as the length and width of the floral tube, as well as the placement of nectaries are common features among the plants. The flying characteristics and mouthparts are convergent among the pollinating insects, butterflies, moths, bats, and birds. It is surprisingly easy to find functional similarities among distantly related organisms.

B. Groupings among Sessile Organisms

Convergent morphology and anatomy is common among sessile organisms on land and in the sea. I briefly consider terrestrial plant communities that have many of the same convergences described previously for animals. Families of cactus-like plants in the Old World (Euphorbiacea) and New World (Cactaceae) dominate arid environments but are unrelated. Much has been written about the remarkably convergent forest and biomes that develop under similar environmental conditions (Huston, 1994). There appear to be relatively few adaptive solutions to common environmental conditions. Often, the suite of characteristics that improve ecological function under specific conditions define functional groupings.

Functional groupings among sessile organisms were studied by terrestrial ecologists, who grouped plants according to their size and the location of their growth and regenerative structures. Raunkiaer (1934) showed that particular functional groups (called "life-forms") persisted and dominated the vegetation under similar environmental conditions. Using this scheme, Arctic regions were clearly distinguished from arid or tropical regions. Similar floras, defined at the functional group level, dominated under similar conditions in different regions even though there were often no species in common. This early focus on the size and placement of growth and regenerative structures was based on the realization that gradients in environmental stress and productivity relative to disturbances are critical to the success of specific life-forms. Plants with protected meristems and perenniating structures persisted in stressful or highly disturbed habitats. Where environmental limitations were relaxed, plant size and competitive ability were more important to the flora.

Recently, the field has become popular among marine ecologists and paleontologists who study sessile marine organisms. Again, morphological and functional convergences are viewed more as the rule than the exception. Several growth forms have evolved convergently among widely disparate groups, such as marine algae, Cnidaria, and Bryozoa (Fig. 7). Most extant and extinct species within those higher taxa can be placed within the groups illustrated in Fig. 7. The growth forms represent a morphological progression. The simplest growth form is solitary and can be a single cell as is the case of a diatom or a solitary organism. This group is often composed of small organisms. The next step in a morphological progression of growth forms simply involves a linear series of the single cells or modules, creating simple filaments. These can grow prostrate along the substrate as recumbent algal filaments or as runners or vines among sessile invertebrates. Similar morphologies growing vertically escape the benthic boundary layer. They experience greater water flow, and most of the cells of these groups are in contact with their aqueous environment. The more massive multiserial or multiseriate organisms allow three-dimensional space to be exploited. Several ecological processes, such as productivity and competition, vary predictably as sessile organisms become larger and more massive. With increasing mass, fewer cells directly contact the marine environment. Fewer still are in contact with the environment among the encrusting forms because half of their surface area is affixed to primary substrate. The growth forms in Fig. 7 exhibit progressively increasing size and generally decreasing surface area to volume ratio.

For benthic marine algae of all three major divisions (= phyla), the progression from single-celled microalgae to filamentous, foliose, larger erect forms and finally to crustose algae (Fig. 7) corresponds to a reduction in productivity per unit biomass (Fig. 8A; Steneck and Dethier, 1994). In general, microalgae and filamentous forms with a high surface area to volume ratio are capable of the highest mass-specific productivity. This decreases progressively from filamentous forms to the more massive erect forms including kelp and is lowest

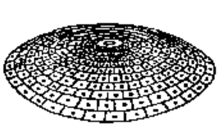
Growth Forms	Construction & Characteristics	Examples		
		Algal	Coral	Bryozoans
	Solitary individual or Single cell (high SA/V) Small size	Microalgae Diatoms	Solitary *Flabellum* *Fungia*	
	uniserial colony or uniseriate prostrate filament	Recumbant filaments (turf) *Herposiphonia*	"Runners & Vines"	*Pyripora*
	uniserial colony or uniseriate or oligoseriate erect filament	Cladophora, Sphacelaria, Polysiphonia	*Madrepora*	*Scruparia*
	Multiserial colony or multiseriate, corticated erect, large size (exploits vertical space)	Corticated and Leathery macroalgae *Gracilaria, Sagassum* kelp (Laminariales)	Branchers *Pocilopora,* *Acropora*	*Cystisella*
	Multiserial or multiseriate encrusting (Lowest SA/V), expansive but low profile (exploits horizontal space)	Crustose *Lithothamnion* *Petrocelis*	*Montipora*	*Stylopoma*

FIGURE 7 Growth forms of sessile marine organisms. Examples include several divisions of marine algae and Cnidaria and Bryozoa phyla. Other groups having some of the same growth forms include the extinct tabulate and rugose corals and modern sponges (after Coates and Jackson, 1985).

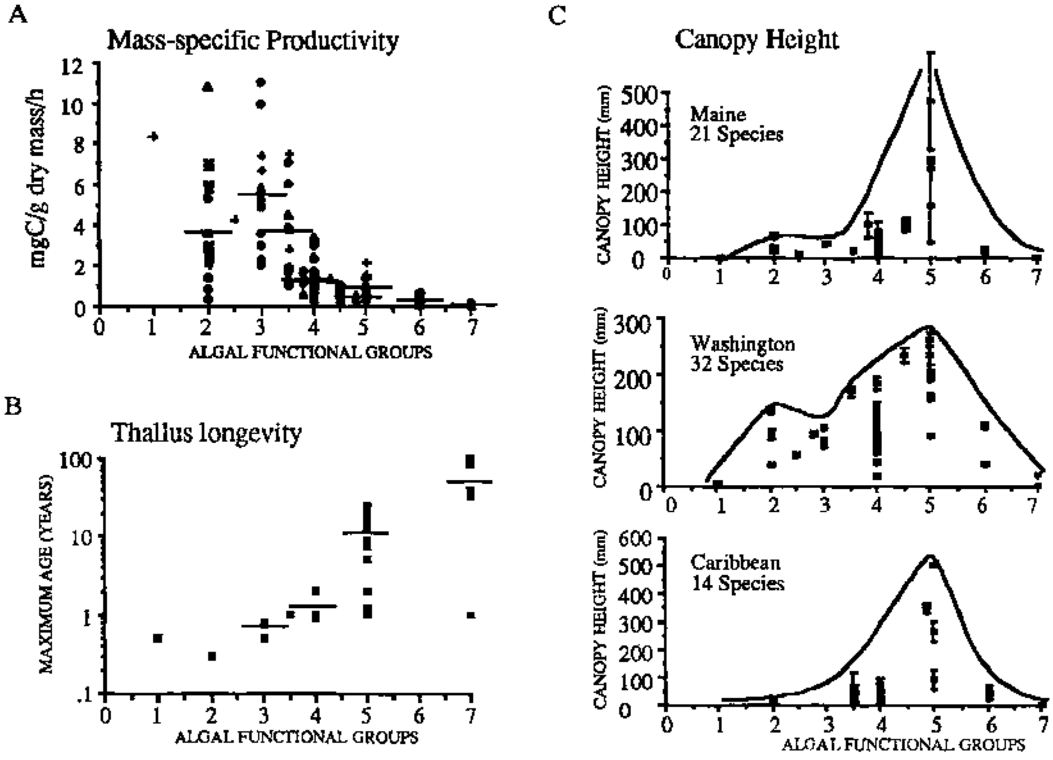


FIGURE 8 Emergent ecological properties of benthic marine algal functional groups. (A) Mass-specific productivity of algae from southern California, the Caribbean, and Hawaii. (B) Thallus longevity of 27 species from 25 published studies. (C) Canopy heights measured in the field for dominant species in Maine, Washington, and the Caribbean. Lines envelop the maximum canopy heights recorded for each algal functional group and points represent average canopy heights per species. Error bars represent one standard deviation (after Steneck and Dethier, 1994).

among the crustose forms. Thallus longevity (Fig. 8B) increases along the same continuum of growth forms. Canopy height, however, shows somewhat of a bell-shaped curve, with the tallest algal forms such as kelp and fucoid algae having intermediate levels of productivity and longevity (Fig. 8C). These emergent ecological properties were shown for cold-water floras of the Atlantic and Pacific as well as for tropical floras of the Caribbean (Fig. 8).

Morphologically and functionally similar organisms have evolved similar ways to overcome similar problems. For example, as organisms become more massive, fewer cells are in direct contact with the marine environment. That is, they have a plumbing problem. Sessile organisms such as plants, algae, and invertebrates evolved physiological integration or a degree of interdependence and cooperation among cells within an organism or modules within a colony. Often this involves sharing cytoplasm or photoassimilates within the organisms and among modules of a colony.

Physiologically integrated colonial organisms have an advantage over physiologically isolated aclonal organisms by sharing cytoplasm among members of the colony. Sessile organisms can respond from a stimulus by growing at a place distant from the stimulus. Wound healing can be facilitated and mortality of the genet reduced by physiological integration. Physiological integration evolved iteratively among sessile organisms on land and in the sea. This allows land plants such as vines to colonize the inhospitable habitats of sand dunes or tree canopies by being physiologically integrated. The most abundant marine algae, the red encrusting coralline algae, are integrated and this is thought to have contributed to their ecological and evolutionary success over unintegrated or poorly integrated algae. Arguably, physiological integration in algae is essential for their morphological diversity. The least physiologically integrated group, Chlorophyta or green algae, have the lowest morphological diversity despite having the same long period of evolutionary history dating back to the Precambrian.

Physiological integration may be a prerequisite for the development of coral reef ecosystems. It evolved independently among cnidarian reef builders: the tabulate, rugose, and finally the scleractinian corals. In each case, the proportion of species with a high degree of physiological integration increased during times of significant reef-building activity. Coral reef ecosystems require high rates of coral growth to keep up or catch up with rising sea level. Reef-building corals today are physiologically plants, with endosymbiotic algae producing carbohydrates necessary for growth. Transport of photosynthetic products throughout the coral colony requires a high degree of physiological integration. Without high rates of coral growth translating to high rates of reef growth, coral reef ecosystems would stand little chance of remaining in the shallow turbulent zone in which they thrive. In fact, it is likely that many coral reef ecosystems that have dominated shallow seas for much of the past 600 million years require the framework-building corals to be physiologically integrated.

Physiological integration must be functionally important to sessile organisms because it evolved independently many times. It is necessary for the similar functional groups found among the corals, algae, and bryozoans (Fig. 7). Colonial integration, or the degree of interdependence and cooperation among modules within a colony, is important for all these groups because it increases colonial function. It also allows the development of nonfeeding defensive or reproductive functions within the colony or clone, and it allows greater structural integrity and morphological diversity. The important functional advantages of physiological integration may account for why it has persisted and become ubiquitous among so many unrelated sessile, colonial, and clonal organisms.

V. THE STRUCTURE AND FUNCTIONING OF NATURAL COMMUNITIES AND ECOSYSTEMS

There exists a broad spectrum of ways to describe patterns. Historically, the emphasis in ecology has been on species. As Naeem (1998) pointed out, species are often viewed as being phyletically "singular" and likely to be "autecologically singular." This was the basis of the important works by Hutchinson and MacArthur (1959), who stressed the uniqueness of species. That no two species can occupy the same ecological niche was the foundation of decades of studies on the structure and function of natural communities and ecosystems. Functional group approaches are different. They stress similarities among unrelated species that share critical organismal features. Convergent evolution in many cases results in phyletically distant organisms sharing ecological properties. The utility of using shared species characteristics for describing communities rests on the idea that in any given environment there are relatively few species attributes important to the structure and functioning of natural communities and ecosystems. These attributes are largely independent of the biodiversity of the ecosystem.

The importance of any species or group can relate to its abundance (measured as population densities, biomass, areal cover, or structural height) or its impact on other species. Whittaker (1965) argued that species' importance should be measured by its productivity or effect on productivity. Obviously, the production of biomass must be approximately matched to its degradation or "disturbance" (*sensu* Grime, 1981). In some systems, species importance depends more on the role of disturbance (e.g., predation) than on productivity. For both disturbance and productivity, properties of the organisms and properties of the environment must be considered simultaneously and independently. Holdridge's (1947) classic "life zone system" for classifying terrestrial vegetation considered only precipitation and temperature as driving productivity and thus creating discrete vegetative communities and ecosystems. Others have shown how disturbances such as storms and fires additionally shape terrestrial plant communities. Although debates continue regarding the primacy of bottom-up or productivity-driven vs top-down or disturbance-driven structure of natural communities, in most systems both components are viewed as important. Therefore, the debate is largely over relative importance.

Productivity and disturbance are the two variables most often identified as environmental structuring agents of natural communities and ecosystems. Huston (1994) reviewed the work of others to develop a model that compared the frequency of disturbance (defined as patch or gap formation) as a function of "potential plant productivity." Relative to these two axes, he showed consistent patterns in the terrestrial realm in species diversity, patch formation, dominance, and competitive ability. General models of community dominants have been advanced by Grime (1981), Southwood (1988), and Steneck and Dethier (1994) (Fig. 9). Specifically, Grime (1981) suggested that, in environments with a high productivity potential (low stress) and low disturbance potential, large, canopy-forming trees will dominate the system (called a "competitive strategy"). Under conditions that favor high productiv-

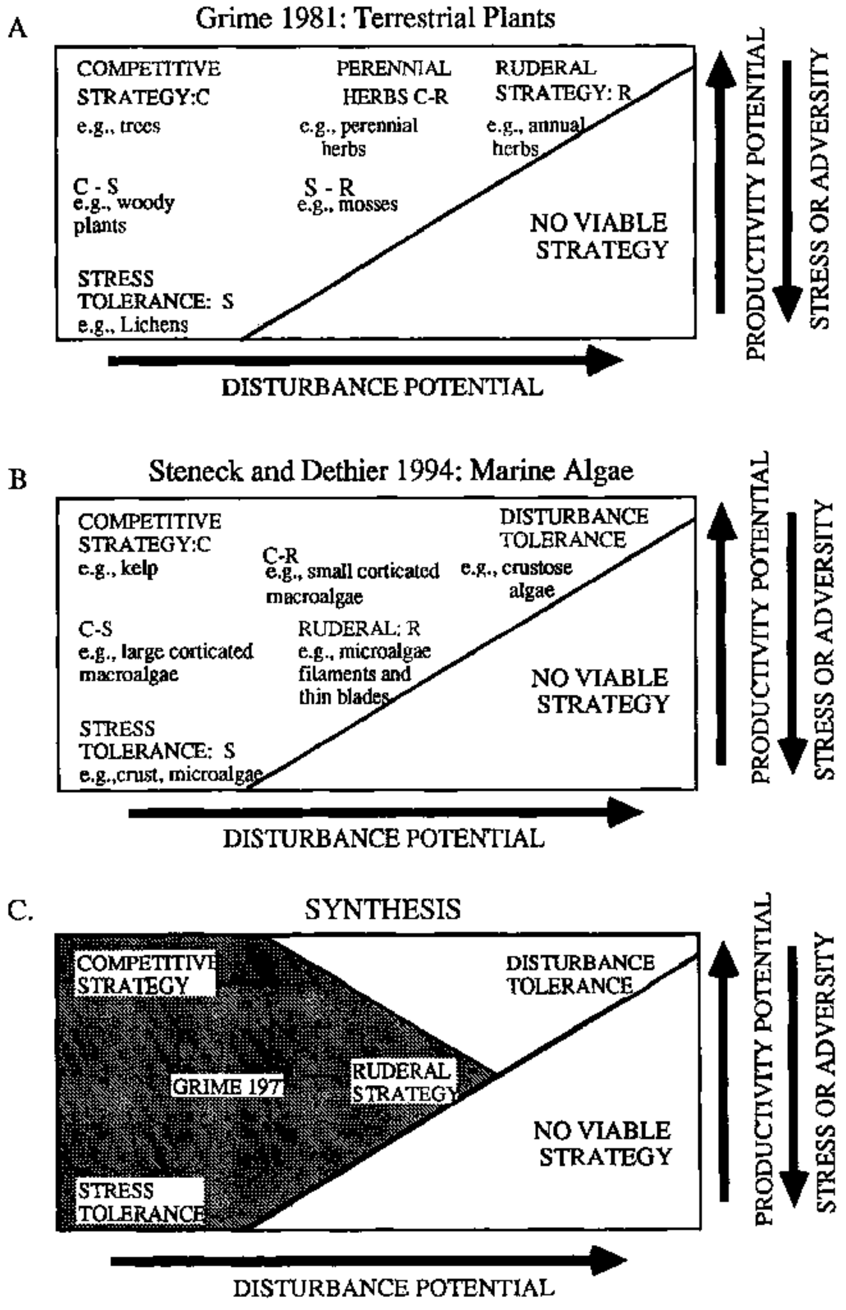


FIGURE 9 Generalized model of community dominants that compares Grime's (1981) model for terrestrial plants (A) with that of Steneck and Dethier (1994) for marine algae (B). (C) A synthesis of A and B. Grime's primary "strategies" or groupings of "competitive," "ruderal," and "stress tolerance" are denoted by the letters C, R, and S, respectively.

ity with increasing productivity, perennial herbs and finally weedy annual herbs will come to dominate the system (called a "ruderal strategy"). Under low disturbance and low productivity potential environments, stress-resistant groups such as lichens will dominate the system (Fig. 9A).

Several studies applied Grime's (1981) model to the marine realm (Fig. 9C). Strong similarities between the two realms surfaced, including that the highest biomass and largest canopy-forming organisms (trees and kelp) grow under high productivity and low disturbance potentials (Fig. 9C). Furthermore, low biomass and canopy height dominated by encrusting growth forms are characteristic of stress-tolerant assemblages in both realms. A difference might occur among the organisms that dominate under conditions of chronically high levels of disturbance. Grime (1981) assumed that it is impossible to withstand such conditions and thus weedy, "ruderal strategies" would persist, whereas Steneck and Dethier (1994) argued that an additional group of organisms tolerant of such disturbances (the coralline algae) predictably dominate such systems. Although it is common to find disturbance-resistant plants under conditions of high stress in the terrestrial realm, they can be found under conditions of low stress (high productivity potential) in the marine realm. It is unclear whether this functional difference between the marine and terrestrial realms is real or perceived. Nevertheless, one value to a functional group approach is that this dialog can occur at all.

VI. ECOSYSTEM DIVERSITY, STABILITY, AND REDUNDANCIES: EXAMPLES FROM CORAL REEFS

Much has been written on this topic, but most of the literature draws from terrestrial studies such as O'Neill *et al.* (1986), Schulze and Mooney (1994a), and Huston (1994). However, rather than reiterating these studies and their myriad examples, I will use coral reef ecosystems as examples in this section. Coral reefs provide excellent, albeit often overlooked, examples. They are spatially discrete and smaller in vertical and horizontal structure than most of their terrestrial counterparts. Most of their functional components are readily visible (there are few below-surface functions), and major alterations and resilience can be measured in years and decades rather than centuries for observing ecosystem changes. Biogeographically, they vary widely in species diversity and thus are ideal for comparing ecosystem structure and function relative to their biodiversity.

If relatively few environmental factors can structure communities and ecosystems at the functional group level, then what role does species diversity play? Here, I consider biodiversity and its role in the stability, structure, and functioning of coral reef ecosystems. If functional groups are composed of ecologically equivalent species, then perhaps diversity within these groups contributes to the redundancy of important functions and thus may be ecosystem insurance for stability.

A. Diversity and Stability

Ecosystem and community stability was once thought to correspond directly with species diversity. MacArthur (1955) theorized that the stability of a community will increase as the number of links in its food web increases. It followed that with the loss of each species "the integrity of the biosphere will degrade in a small but significant way" (Lawton and Brown, 1994, page 255). Alternatively, species richness may be relatively unimportant as long as structuring processes controlled by primary producers, consumers, and decomposers are maintained and function well with very few species. In fact, it is now widely believed that "ecosystem processes often have considerable redundancy built into them" (Lawton and Brown, 1994, page 266). Functional groups can be composed of the redundant species that perform essential ecosystem processes.

There is good empirical evidence that diverse ecosystems are not stable. Coral reefs are among the most diverse and most productive ecosystems in the world, but they are not characteristically stable. In the Caribbean and Indo-Pacific reef coral cover, abundance and dominance suddenly changed as the result of changes in abundance of a single echinoderm species in each of the two regions. In 1983, the mass mortality of the Caribbean sea urchin, *Diadema antillarum*, caused a rapid "phase shift" that increased macroalgal biomass and contributed to the decline of live coral cover in Curacao, St. John, Jamaica, and St. Croix. In all cases, algal assemblage dominance shifted from turf algae (Fig. 8A, Nos. 1–3) to macroalgae (Fig. 8A, Nos. 4–6), and this caused a decline in mass-specific primary productivity. It also precipitated changes in the abundance of other herbivores. The abundance of all reef-building organisms (corals, hydroids, and coralline algae) declined sharply in shallow zones where carbonate production is normally highest.

In the Indo-Pacific, periodic, localized outbreaks of the predatory sea star *Acanthaster planci* caused coral cover declines of all species of between 85 and 94% as reviewed by Connell (1997). Other disturbances, such as storms, sedimentation, toxins, and fishing, have also been shown to cause significant reductions in coral cover. In the cases of very rapid ("acute or pulse") mortality from sea star predation, rapid recovery at decadal timescales was observed. Other longer lasting ("chronic or press") mortality events had slower rates of recovery. However, in the scale of years to decades these highly diverse ecosystems were demonstrably unstable.

Ecosystem resilience is sometimes considered a correlate of stability. Rapid recovery from disturbances was more commonly observed in reefs of the Indo-Pacific than in reefs of the Caribbean, where no recoveries have been documented. Reef recovery relates more to the nature and duration of the disturbances than to the biodiversity of the system. One of the best examples of recovery from collapse is in some reefs of Hawaii, although the species diversity there was as low as that found throughout the Caribbean (Table I).

B. Diversity and Ecosystem Structure

The basic structure of coral reef ecosystems does not vary with biodiversity. Coral reefs throughout the Indo-Pacific, Caribbean, and at Clipperton Atoll (eastern Pacific) have similar morphological zonation of corals despite a two orders of magnitude difference in coral species richness (Table I). Branching corals dominate shallow zones, mound-shaped forms at greater depths, and platy corals at still greater depth. At Clipperton Atoll, one of the most species-depauperate reefs in the world, Glynn and coworkers (1996) measured vertical growth rates of coral similar to those found for similar morphologies at similar depths elsewhere in the Indo-Pacific. Steneck (1988) found similar patterns of zonation, distribution, and abundance among marine algae and herbivores at the functional group level in most studied reef systems of the Indo-Pacific and Caribbean. Others showed that functional characteristics of reef fish, such as size, activity rhythms, and relative abundances among ecological and trophic categories, were similar regardless of the species richness of the fish fauna.

C. Diversity and Ecosystem Function

Species diversity is always low at high trophic levels. Often a single, "apex" predator is critical to the structure of some systems. However, at lower trophic levels, diversity increases, and it is at these levels (among herbivores and primary producers) that the relationship between ecosystem diversity and function is best examined at the level of functional groups.

1. Herbivores

Reef-dwelling herbivores have been grouped according to their grazing abilities on algal communities (Table II). Deep grazing (i.e., carbonate scraping herbivores), parrotfishes, and certain sea urchins (e.g., *Diadema*) have the greatest impact on algal abundance. Other fishes, such as surgeonfish (acanthurids), certain damselfishes and blennies, and some gastropods, are "denuding" herbivores in that they can significantly reduce fleshy algal biomass when at sufficiently high densities. This group does not bite into limestone structures of reefs. They cannot feed on crustose corallines and have a limited ability to consume articulated algae and large leathery macroalgae. Nondenuding herbivores have little or no ability to reduce algal biomass and include territorial damselfishes and smaller herbivores, such as amphipods and polychaetes. Each of these herbivore functional groups performs distinctly different ecological roles in the ecosystem, and all were found on the species-depauperate Clipperton Atoll. The abundance of species within a functional group compensates for low diversity within the group. For example, high rates of scraping herbivory were evident at Clipperton, but all such grazing resulted from a single parrotfish species and several species of diadematid urchins. The abundant crustose coralline algae, which are often found on reefs experiencing high rates of grazing, were covered with graze marks from scraping herbivores (Table II). High rates of grazing, low fleshy algal biomass, and a high abundance of both corals and coralline algae are characteristics of relatively pristine reefs. That these conditions were found in as species depauperate a system as Clipperton Atoll suggests that ecological roles

TABLE I

Comparison of Coral Reef Biodiversity[a]

Region	Coral species	Reference	Fish species	Reference
Indo-Pacific	350	Karlson and Cornell (1998)	3000	Lieske and Myers (1994)
Caribbean	50	Karlson and Cornell (1998)	750	Lieske and Myers (1994)
Hawaii	50	Hoover (1998)	680	Hoover (1993)
Clipperton (Eastern Pacific)	7	Glynn *et al.* (1996)	101	Robertson and Allen (1996)

[a] Species richness of scleractinian corals and reef fishes.

TABLE II

Functional Groups of Herbivores Common on Coral Reefs[a]

Functional group	Specific group	Common reef-dwelling genera
Scraping	Parrotfish	*Scarus, Sparisoma*
	Sea urchins	*Diadema, Echinometra, Echinothrix, Lytechinus*
	Limpets, chitons	*Acmaea, Acanthopleura, Acanthochitona, Choneplax*
Denuding	Surgeonfish, signids,	*Acanthurus, Naso, Zebrasoma, Siganus*
	Few pomacentrids	*Microspathodon*
	Blennies, kyphosids	*Ophioblennius*
Nondenuding	Damselfish	*Stegastes, Pomacentrus, Eupomacentrus*
	Snails (excluding limpets)	*Nerita, Cittarium, Tectarius, Astraea, Fissurella*
	Amphipods	*Amphithoe*, tanaids
	Polychaetes	Eunicids, syllids.

[a] Based on more than 30 published studies (after Steneck, 1988).

of functional groups can be independent of species diversity.

2. Primary Producers

Primary productivity is a significant component of ecosystem functioning. A functional approach to ecosystem study often "implies that energy flow and nutrient cycling are somehow more important or more fundamental than the biotic entities performing the function" (O'Neill *et al.*, 1986, page 10). Although there may be no species in common, the same functional groups of marine algae are found on virtually all coral reefs. Since the productive components of reefs (algal turfs, corals, and macroalgae) all occur in approximately similar proportions, ecosystem-level productivity is fairly consistent. Thus, Caribbean reefs with only a fraction of the species richness of the western Indo-Pacific reefs have similar rates of gross primary productivity. Other factors, such as the geological history of the reefs and their proximity to land, have much greater impacts on ecosystem production than do reef biogeography or species diversity per se. Note that these patterns of consistent rates of productivity among unstressed reefs do not negate the well-known pattern of low diversity in stressed habitats.

D. Equivalent Species and Ecosystem Redundancy

Species within a function group, by definition, are ecologically equivalent and as such provide a degree of redundancy to the ecosystem. One test of functionally equivalent species is if their abundance compensates as other species within the functional group change. Morphological and functional equivalency is evident within turf algae on coral reefs. Many genera of different divisions look similar. They often grow as prostrate branching filaments, with relatively few laterally divided "pericentral" cells. Many species among these genera (Fig. 10A) have about the same high rates of productivity (Fig. 8A) and they respond similarly to injuries. In the fore reef of St. Croix, dominance and abundance of algal species changed dramatically from season to season at the species level but very little at the functional group level (Fig. 10B). That algal community biomass remained relatively constant throughout the year, as species composition fluctuated greatly, suggests that species within functional groups compensate for species fluctuations.

It is necessary to sort out functionally important changes from redundant species "noise." On the fore reef of St. Croix (Fig. 10), seasonal changes among species were so great that it was impossible to describe it as having a characteristic dominant species. However, at the functional group level, a clear pattern of filamentous and microalgal (i.e., turf algal assemblage) dominance was evident. These algae readily coexist with reef-building corals. In contrast, macroalgae comprised a minor component of this flora and appeared only in winter and spring. The functional distinction between macroalgae and the dominant micro- and filamentous algae is important. The algal turf assemblage was composed of species which have the [illegible] mass-specific productivity (i.e., algal functional [illegible], [illegible]s. 1 and 2 in Fig. 8A). Trophically, these alg[illegible] [illegible]ies support very high levels of grazing. Th[illegible] [illegible]us fish and urchins cannot distinguish sp[illegible] [illegible] [illegible]e diverse algal turfs, and thus no species [illegible] possible

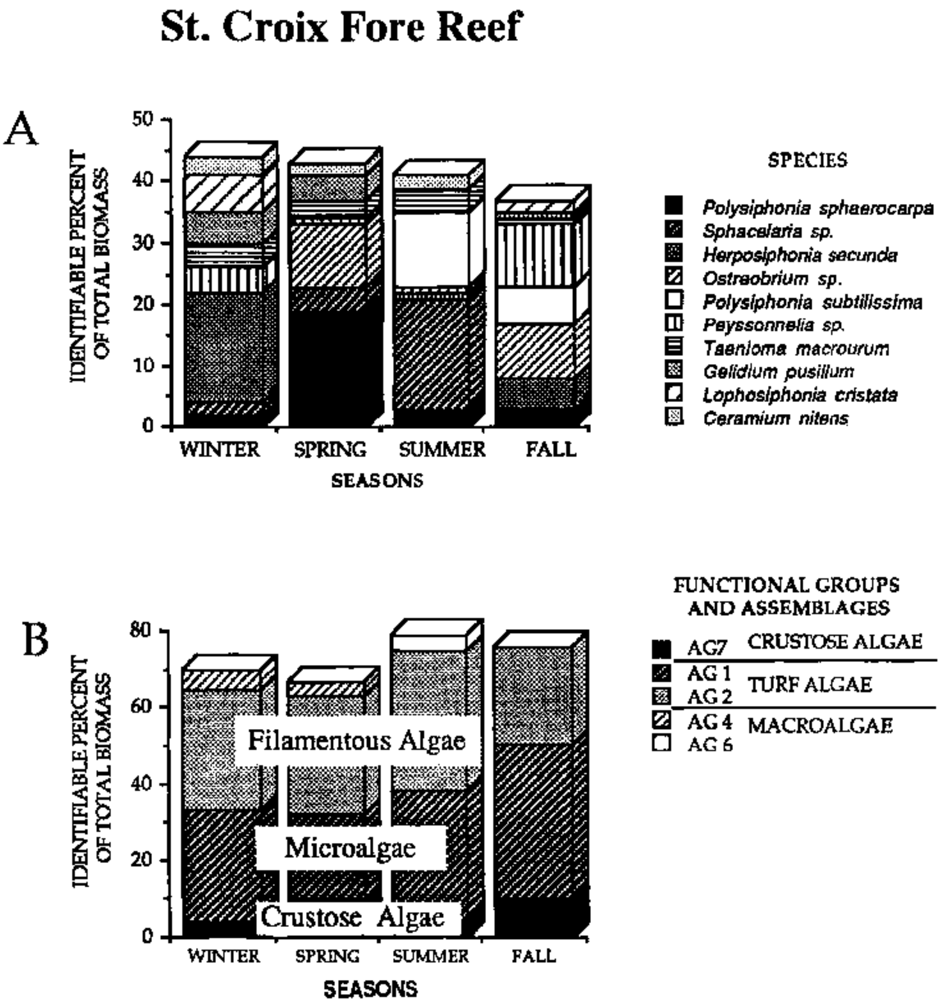


FIGURE 10 Seasonal stability of the 10 most abundant algal species (A) and functional groups (B) on a fore reef of St. Croix. Differences in the total percentage represented among seasons for species (A) and functional groups (B) resulted from the different proportions of unidentifiable components when samples were analyzed at species vs functional group levels. Functional group numbers correspond to the ones plotted in Fig. 8. Functional groups were further combined into common tropical algal assemblages (i.e., crustose algae, turf algae, and macroalgae) (after Steneck and Dethier, 1994).

and there are no know chemical deterrents within this assemblage. Any of the filamentous algal species could have gone locally extinct (many did seasonally) without functionally changing the system. As such, many of the species within these functional groups are redundant.

Compensation is also evident among ecologically equivalent functional groups of herbivores. In shallow reef zones, the grazing effects from parrotfish and sea urchins are functionally similar (Table II). They are both capable of grazing deeply into limestone substrates and can control the structure of algal communities on reefs. Several studies have reported that deep-grazing sea urchins and fish compete and show compensatory changes in relative abundance when one group is excluded or removed. Hay (1984) found that reefs where fishing was important were dominated by sea urchin grazing, whereas with little fishing pressure reefs were dominated by herbivorous fishes (e.g., scraping and denuding fishes in Table II). Subsequent studies at several locations in the Caribbean and Indian Ocean found that when urchins were removed experimentally or naturally due to a disease-induced mass mortality, fish abundance and grazing increased as urchin abundance declined. In regions of the Caribbean traditionally having a relatively low abundance of *Diadema antillarum*, another urchin, *Echinometra viridis*, was abundant and filled *Diadema's* functional role.

E. Ecosystem Insurance

Species diversity within functional groups buffers ecosystems against species loss, but often there is "a threshold of change that will overwhelm the damping effect of biodiversity, with an associated break point of ecosystem function to quite different levels" (Schulze and Mooney, 1994b, page 501). Unfished reefs contain high within-functional group diversity at several levels. There are several species of large carnivores (e.g., groupers and snappers), denuding herbivores (e.g., surgeonfish), and scraping herbivores (e.g., parrotfish) that, when abundant on reefs, control the reef's overall structure and functioning. Fishing affects reefs by first removing large-bodied carnivores and then herbivores. Reefs under particularly intense fishing, such as in Jamaica, are left with few or very small fishes. Compensatory population increases in the sea urchin *D. antillarum* probably occurred (Fig. 11), causing it to become hyperabundant and the prime grazer of the reefs of Jamaica and elsewhere in the Caribbean. Levitan (1992) showed that *Diadema's* increasing role as a grazer throughout the Caribbean corresponded with increasing human populations and associated fishing pressure. On many Caribbean reefs this herbivore became the primary grazer within the functional group of scraping herbivores (Table II). Arguably, the diversity within this functional group had decreased to one species (Fig. 11). When *D. antillarum* suffered a mass mortality in 1983, the "break point" had been exceeded and reef ecosystems throughout the Caribbean collapsed.

F. Keystone Species to Functional Groups: Some Species Are More Equal Than Others

Some species have a disproportionately great impact on ecosystems. Hurlbert (1997) noted, "The general functional importance of a species [can be] defined as the sum, over all species, of the changes (sign ignored) in productivity which would occur on removal of the particular species from the [system]" (page 369). Most communities and ecosystems are dominated by a few species that are both abundant and important to that system (Fig. 12). The few tree species that dominate given forests, corals that dominate zones of reefs, and specific grasses that comprise specific prairies need not be functionally grouped to be easily understood. The loss of community or ecosystem dominants has a large impact on natural systems. Diseases can eliminate dom-

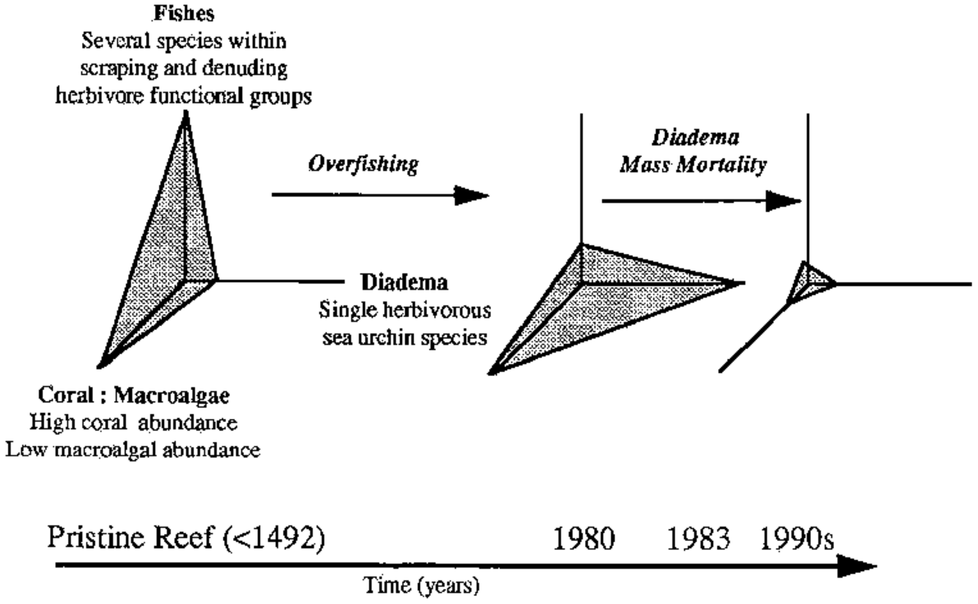


FIGURE 11 Temporal changes in Jamaica's coral reef ecosystem in coral to macroalgal abundance relative to herbivorous fish and sea urchin, *Diadema antillarum*. After herbivorous fishes were extirpated by overfishing, grazing remained high due to compensatory increases in *Diadema* abundance. The reef ecosystem collapsed soon after the mass mortality of the urchin (see text) (adapted with permission from Done, Ecological criteria for evaluating coral reefs and their implications for managers and researchers, 183–192. © 1995 Springer-Verlag).

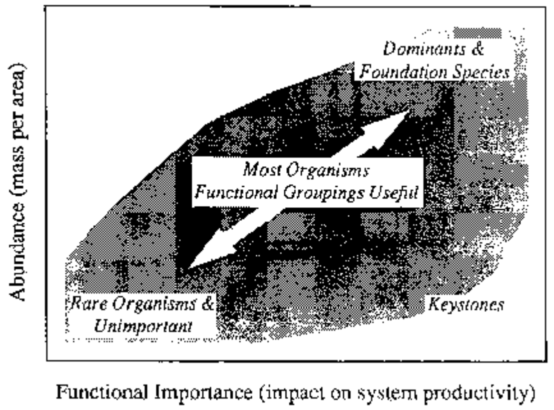


FIGURE 12 Functional importance and abundance of species. The relatively low diversity of dominant/foundation and keystone species as well as the relatively low ecosystem importance of rare organisms make these groups poorer candidates for functional groupings (modified from Hurlbert, 1997).

inant organism with impacts registering throughout the system. American chestnut blight eliminated chestnuts from forests in eastern North America. White-band disease killed most of the dominant elkhorn and staghorn corals and a pathogen caused the *Diadema* decline in shallow Caribbean reefs. Rinderpest decimated wildebeest populations in Africa, and paramoebae caused a mass mortality of the green sea urchin in Nova Scotia. Each of these dominant or "foundation" (*sensu* Dayton, 1972) species underwent a radical decrease in abundance, causing their respective ecosystems to change. In these cases, the functional importance of these species is unquestioned.

Keystone species have been redefined by Paine (1995) and Power *et al.* (1996) to include organisms that are functionally important but in relatively low abundance in the system. Pathogens aside, this limits keystone species to apex predators. Classic keystone predators such as sea otters or the predatory sea star, *Pisaster*, are the only species in their system and thus define their ecological function. However, many systems do not have an obvious apex predator. In contrast with the marine systems of the eastern North Pacific in which sea otters and *Pisaster* roam, the western North Atlantic has no single apex predator. Top coastal predators include cod, haddock, hake, and wolffish. When all these top predators were extirpated simultaneously by overfishing, the system changed. Therefore, whether there is a single apex predator (i.e., a keystone species) or a functional group filling the same role matters little. In other words, the ecological function can be carried out by a single species or by a suite of species, but the ecosystem will not change unless the ecological function changes significantly. In the end, the ecological function, independent of the diversity, may be most important to the structure of ecosystems.

The utility of functional groupings relates directly to the diversity of the system, community, or trophic level of interest. It will be most useful in sorting out the functional importance in diverse systems among moderate to very abundant organism (i.e., the center of Fig. 12). Functional groups lose their utility among the rare, dominant, and keystone species (i.e., the periphery of Fig. 12). Not only is species diversity low at high trophic levels but also it is characteristically low in certain stressed habitats, such as deserts, the rocky intertidal zone, estuaries, and high-altitude or high-latitude habitats. Functional group utility is likely to be greatest in diverse systems in which many species perform similar ecological functions.

VII. CONCLUSIONS

Despite Earth's remarkable biodiversity, the range of functionally different organisms is surprisingly low. Phyletic constraints on evolution relative to biomechanical limits of what works in nature limit the number of functional groups in any system to a management level. In highly diverse ecosystems there are many redundant species within functional groups. These redundancies provide a buffer against ecosystem collapse should individual species within the group become rare or extinct. Ultimately, functional groups may provide a low-resolution tool for accurately predicting ecosystem change. Research to improve how we identify and define functional groups is likely to increase our understanding of the structure and functioning of natural communities and ecosystems.

Acknowledgments

Drafts were read by E. Annis, A. Leland, J. Vavrinec, and S. Zimsen, with additional help from P. Yund and L. Watling. Greg Welch helped with manuscript preparation. To all, I am grateful.

See Also the Following Articles

BIODIVERSITY GENERATION, OVERVIEW • CLADOGENESIS • COEVOLUTION • GUILDS • MAMMALS, BIODIVERSITY OF • REEF ECOSYSTEMS

Bibliography

Bambach, R. K. (1985). Classes and adaptive variety: The ecology of diversification in marine faunas through the Phanerozoic. In

Phanerozoic Diversity Patterns: Profiles in Macroevolution (J. W. Valentine, Ed.), pp. 191–253. Princeton Univ. Press, Princeton, NJ.

Begon, M., Harper, J. L., and Townsend, C. R. (1986). *Ecology, Individuals, Populations and Communities*. Sinauer, Sunderland, MA.

Coates, A. G., and Jackson, J. B. C. (1985). Morphological themes in the evolution of clonal and aclonal invertebrates. In *Population Biology and Evolution of Clonal Organisms* (J. B. C. Jackson, L. W. Buss, and R. E. Cook, Eds.), pp. 67–106. Yale Univ. Press, New Haven, CT.

Connell, J. H. (1997). Disturbance and recovery of coral assemblages. *Coral Reefs* 16 (Suppl.), S101–S113.

Crimes, T. P., and Harper, J. C. (Eds.) (1977). *Trace Fossils 2, Geological Journal* Special Issue No. 9. Seel House Press, Liverpool, UK.

Cunningham, C. W., Blackstone, N. W., and Buss, L. W. (1992). Evolution of king crabs from hermit crab ancestors. *Nature* 355, 539–542.

Darwin, C. (1859). *The Origin of Species by Means of Natural Selection*. Random House, New York. (Reprinted by Modern Library)

Done, T. J. (1995). Ecological criteria for evaluating coral reefs and their implications for managers and researchers. *Coral Reefs* 14, 183–192.

Glynn, P. W., Veron, J. E. N., and Wellington, G. M. (1996). Clipperton Atoll (eastern Pacific): Oceanography, geomorphology, reef-building coral ecology and biogeography. *Coral Reefs* 15, 71–99.

Grime, J. P. (1981). *Plant Strategies and Vegetation Processes*. Wiley, New York.

Hay, M. E. (1984). Patterns of fish and urchin grazing on Caribbean coral reefs: Are previous results typical? *Ecology* 65, 446–454.

Holdridge, L. R. (1947). Determination of world plant formations from simple climatic data. *Science* 105, 367–368.

Hoover, J. P. (1993). *Hawaii's Fishes: A Guide for Snorkelers, Divers, and Aquarists*. Mutual, Honolulu, HI.

Hoover, J. P. (1998). *Hawaii's Sea Creatures: A Guide to Hawaii's Marine Invertebrates*. Mutual, Honolulu, HI.

Hurlbert, S. H. (1997). Functional importance vs keystoneness: Reformulating some questions in theoretical biocenology. *Austr. J. Ecol.* 22, 369–382.

Huston, M. A. (1994). *Biological Diversity. The Coexistence of Species on Changing Landscapes*. Cambridge Univ. Press, Cambridge, UK.

Hutchinson, G. E., and MacArthur, R. (1959). A theoretical ecological model of size distributions among species of animals. *Am. Nat.* 93, 117–126.

Karlson, R. H., and Cornell, H. V. (1998). Scale-dependent variations in local vs regional effects on coral species richness. *Ecol. Monogr.* 68, 259–274.

Lawton, J. H., and Brown, V. K. (1994). Redundancy in ecosystems. In *Biodiversity and Ecosystem Function* (E. D. Schulze and H. A. Mooney, Eds.), pp. 255–268. Springer, New York.

Levitan, D. R. (1992). Community structure in times past: Influence of human fishing pressure on algal urchin interactions. *Ecology* 73, 1597–1605.

Lieske, E., and Myers, R. (1994). *Coral Reef Fishes: Caribbean, Indian Ocean, and Pacific Ocean Including the Red Sea*. Princeton Univ. Press, Princeton, NJ.

MacArthur, R. H. (1955). Fluctuations of animal populations and a measure of community stability. *Ecology* 35, 533–536.

Moody, K. E., and Steneck, R. S. (1993). Mechanisms of predation among large decapod crustaceans of the Gulf of Maine coast: Functional vs phylogenetic patterns. *J. Exper. Marine Biol. Ecol.* 168, 111–124.

Naeem, S. (1998). Species redundancy and ecosystem reliability. *Conserv. Biol.* 12, 39–45.

Niklas, K. J. (1994). *Plant Allometry: The Scaling of Form and Process*. Univ. of Chicago Press, Chicago.

O'Neill, R. V., DeAngelis, D. L., Waide, J. B., and Allen, T. F. H. (1986). *A Hierarchical Concept of Ecosystems*. Princeton Univ. Press, Princeton, NJ.

Paine, R. T. (1995). A conversation on refining the concept of keystone species. *Conserv. Biol.* 9, 963–964.

Peters, R. H. (1983). *The Ecological Implications of Body Size*. Cambridge Univ. Press, Cambridge, UK.

Power, M. E., Tilman, D., Estes, J. A., Menge, B. A., Bond, W. J., Mills, L. S., Daily, G., Castilla, J. C., Lubchenco, J., and Paine, R. T. (1996). Challenges in the quest for keystones. *BioScience* 46, 609–620.

Raunkiaer, C. (1934). *The Life Forms of Plants and Statistical Plant Geography*. Clarendon, Oxford.

Robertson, D. R., and Allen, G. R. (1996). Zoogeography of the shorefish fauna of Clipperton Atoll. *Coral Reefs* 15, 121–131.

Schulze, E., and Mooney, H. A. (Eds.) (1994a). *Biodiversity and Ecosystem Function*. Springer, New York.

Schulze, E.-D., and Mooney, H. A. (1994b). Ecosystem function of biodiversity: A summary. In *Biodiversity and Ecosystem Function* (E.-D. Schulze and H. A. Mooney, Eds.), pp. 497–510. Springer, New York.

Sepkoski, J. J., Jr. (1984). A kinetic model of Phanerozoic taxonomic diversity. III. Post-Paleozoic families and mass extinctions. *Paleobiology* 10, 246–267.

Southwood, T. R. E. (1988). Tactics, strategies and templets. *Oikos* 52, 3–18.

Steneck, R. S. (1982). A limpet coralline alga association: Adaptations and defenses between a selective herbivore and its prey. *Ecology* 63, 507–522.

Steneck, R. S. (1988). Herbivory on coral reefs: A synthesis. *Proc. 6th Int. Coral Reef Symp. Austr.* 1, 37–49.

Steneck, R. S., and Dethier, M. N. (1994). A functional group approach to the structure of algal-dominated communities. *Oikos* 69, 476–498.

Steneck, R. S., and Watling, L. E. (1982). Feeding capabilities and limitations of herbivorous molluscs: A functional group approach. *Mar. Biol.* 68, 299–319.

Thompson, D'arcy W. (1966). *On Growth and Form*. Cambridge Univ. Press, Cambridge, UK.

FUNGI

Thomas J. Volk[1]
University of Wisconsin–La Crosse

GLOSSARY

ectomycorrhiza (pl. ectomycorrhizae; adj., ectomycorrhizal) Part of a mutualistic relationship between a fungus (usually a basidiomycete, but sometimes an ascomycete) and a host plant in which hyphae aggregate as an extra surface around roots of the plant, aiding nutrient transfer between the plant and fungus.

endomycorrhiza (pl. endomycorrhizae; adj. endomycorrhizal) Part of a mutualistic relationship between a fungus (a zygomycete) and a host plant in which no sheath is formed. Nutrient transfer occurs with the aid of highly branched arbuscules.

exoenzymes Digestive enzymes excreted by fungi into the environment to digest materials externally.

fruiting body Sexual reproductive structure of a fungus.

fungus A member of the kingdom Fungi. Members of this kingdom are heterotrophic (requiring a preformed organic source, i.e., not able to make their own food), eukaryotic, have walls of chitin, and reproduce by means of spores.

karyogamy Fusion of nuclei.

mushroom Vernacular word for a large, fleshy fruiting body consisting of a cap with gills or pores on the undersurface (or sometimes flat or with teeth or folds), usually on a stalk, and producing sexual spores. The term is usually reserved for members of the Basidiomycota.

mutualism Two organisms living together for the mutual benefit of both.

mycelium (pl. mycelia) Vegetative filamentous body of a fungus; mass of hyphae from a single individual.

parasite Organism that obtains its nutrition from a living host, and in so doing harms the host.

plasmogamy Fusion of the cytoplasm of two cells.

spore One- to several-celled propagule of totipotent cytoplasm with cell walls, produced by cell division with concomitant meiosis or mitosis, that may serve for dispersal or overseasoning, but that does not contain an embryo.

substrate Material or host from which a fungus derives its nutrition.

toadstool vernacular name for a poisonous mushroom.

WHEN SOMEONE MENTIONS "FUNGI" you may think immediately of mushrooms on pizza or maybe

---

[1] www.wisc.edu/botany/fungi/volkmyco.html

Encyclopedia of Biodiversity, Volume 3

moldy food in your refrigerator. But in fact fungi are everywhere and affect our lives every day, from edible mushrooms to industrially important products to plant helpers, plant pathogens, and human diseases.

I. INTRODUCTION

Many fungi are detrimental in causing a large number of plant diseases that result in the loss of billions of dollars worth of economic crops each year. They also cause a number of animal diseases, including many human maladies. On the other hand, there is a long and rapidly growing list of useful fungi. Fungi have been used in the preparation of foods and beverages for thousands of years, and there are many mushrooms that are edible. Industry has used other fungi in the manufacture of many valuable organic compounds, including organic acids, vitamins, antibiotics, and hormones. They have been used in the research laboratory to study metabolic pathways, mineral nutrition, genetics, and a variety of other problems. But perhaps their greatest contribution has been, and continues to be, their role in recycling carbon and other essential elements in the ecosystem. Because all fungi are heterotrophic (see Section V), they rely on organic material, either living or dead, as a source of energy. Thus, many are excellent scavengers in nature, breaking down dead animal and vegetable material into simpler compounds that become available to other members of the ecosystem.

"Mycology" comes from the Greek words *mykos*, which means "fungi," and "-logy," which means "the study of." Mycologists study many aspects of the biology of fungi, usually starting with their systematics, taxonomy, and classification, and continuing on to their physiology, ecology, pathology, evolution, genetics, and molecular biology. There are quite a few aspects of applied mycology, such as plant pathology, human pathology, fermentation, mushroom cultivation, and many other fields.

II. WHAT ARE FUNGI?

A. Characteristics of Fungi That Separate Them from the Other Kingdoms

The most significant characteristics of fungi that place them apart from the plant, animal, and other kingdoms are the following:

- Fungi are eukaryotic.
- They are nonvascular organisms, meaning they have no specialized transport tissues.
- Most reproduce by means of spores, usually wind-disseminated, and occasionally by insects in some species.
- Both sexual (meiotic) and asexual (mitotic) spores may be produced, depending on the species and conditions.
- Fungi are typically not motile, although a few (e.g., chytrids) have a motile phase.
- Like plants, sexually reproducing fungi have an alternation of generations, although the generations may be different from those of plants.
- The fungal vegetative body may be unicellular (yeasts) or, more often, composed of microscopic threads called hyphae.
- Fungal cell walls are similar in structure to those of plants but differ in chemical composition—fungi cell walls are composed mostly of chitin, which are β-1,4 linkages of *n*-acetylglucosamine. In contrast, plant cell walls are composed mostly of cellulose, which are β-1,4 linkages of glucose. Many plants also contain lignin in their secondary walls.
- Fungal cytoplasmic ultrastructure is broadly similar to that of plant cells, but differs significantly in the kinds of organelles and their structures.
- Fungi are heterotrophic (i.e., "other feeding" in that they must feed on preformed organic material), not autotrophic (i.e., "self-feeding" in that they make their own food by photosynthesis, like most plants and algae).
- Unlike animals (which are also heterotrophic), which ingest and then digest, fungi first digest and then ingest. Fungi produce exoenzymes to accomplish this.
- Most fungi store their food as glycogen (like animals)—plants store food as starch.
- Fungal cell membranes have a unique sterol, ergosterol, which replaces the cholesterol found in mammalian cell membranes.
- The lysine biosynthesis pathway in fungi is different.
- The microtubule protein formed during nuclear division is different from that of all other organisms.
- Most fungi have very small nuclei, with little repetitive DNA, usually with few chromosomes.
- Mitosis and meiosis are generally accomplished without dissolution of the nuclear envelope.

There are about 70,000 named species of fungi and this is believed to be about 5% of the total number of species that exist in nature. If this is the case, 95% of all fungal species are unknown to science and do not yet have names!

In the outdated two-kingdom system, fungi were included in the plant kingdom, as were almost all walled organisms that are not motile. In almost all systematic schemes today, the heterotrophic and eukaryotic fungi are placed in their own kingdom, cleverly called the kingdom Fungi.

The eukaryotes, organisms with nuclei and membrane-bound organelles, can be divided into five kingdoms: Fungi, Plantae, Animalia, Protista, and Stramenopila. Recent DNA evidence suggests that fungi are more closely related to the animals than to plants. Prokaryotes, without nuclei and membrane-bound organelles, include the Bacteria and the Archaea.

As with all taxonomy, the names of various taxa of fungi each have a specific ending that refers to their taxonomic level. The fungal kingdom has its own endings for many taxa; these include the suffix -myc. It is of considerable advantage to be familiar with these endings so that one knows immediately what taxonomic level is being referred to when confronted with a particular name. The names of fungi are "regulated" by the International Code of Botanical Nomenclature, and thus endings for the order and family levels are the same as those of plants (Table I).

B. Fungal Roles in the Ecosystem

Lack of chlorophyll profoundly affects the lifestyle of fungi. Since they are not dependent on light, they can occupy dark habitats and can grow in any direction. The exoenzymes allow them to invade the interior of a substrate with absorptive filaments. Fungi may gain their nutrition from dead organisms, in which case they are called saprophytes. Some fungi derive their nutrition from living organisms; these are called symbionts. Symbionts can be further divided on the basis of whether or not they harm their host. Parasites cause harm to the host, whereas mutualists engage in a reciprocally beneficial association with their host. These categories are further described in Section V.

TABLE I

Taxonomic Endings for the Fungi

Taxonomic level	Taxonomic ending	White button pizza mushroom	The morel
Domain	Eukarya	Eukarya	Eukarya
Kingdom	Fungi	Fungi	Fungi
Phylum	-mycota	Basidiomycota	Ascomycota
Class	-mycetes	Hymenomycetes	Discomycetes
Order	-ales	Agaricales	Pezizales
Family	-aceae	Agaricaceae	Morchellaceae
Genus	—	*Agaricus*	*Morchella*
Species	—	*Agaricus bisporus*	*Morchella esculenta*

C. Importance of Studying Fungal Biodiversity

Fungi affect human lives in many and varied ways, so it is important to know something about the fungi to be able to control or exploit them for our own purposes. For example, more than 90% of known fungal species have never been screened for antibiotics or other useful compounds. However, even more important is the role that fungi play in the ecosystem. They are a vital part of the links in the food web as decomposers and pathogens and are important in grassland and forest ecosystems alike. Fungi have many different kinds of associations with other organisms, both living and dead. To learn more about the impact that fungi have on our lives, we must learn a lot more about them.

Many fungi are harmful to human interests. They can cause human disease, either directly or through their toxins, including mycotoxins and mushroom poisons. They can also cause diseases of plants and animals (e.g., crops, fruit trees, farm animals). Very often fungi cause rot and contamination of foods—most of us probably have something green and moldy in the back of our refrigerator right now. They can destroy almost every kind of manufactured good, with the exception of some plastics and some pesticides.

On the other hand, many fungi are very useful to humans. Of course, there are many edible mushrooms. Yeasts have been used for baking and brewing for many millennia. Antibiotics such as penicillin and cephalosporin are produced by fungi. The immunosuppressive antirejection transplant drug cyclosporin is produced by the ascomycete *Tolypocladium inflatum*. Steroids and hormones—and even birth control pills—are commercially produced by various fungi. Many organic acids are also commercially produced with fungi, for example, citric acid in cola and other soda pop products is produced by an *Aspergillus* species. Some gourmet cheeses such as Roquefort and other blue cheeses, brie, and camembert are fermented with certain *Penicillium* species. Stone-washed jeans, strange as it sounds, are softened by *Trichoderma* species. There are likely many more potential uses that have not yet been explored.

Fungi are also important experimental organisms. They are easily cultured, occupy little space, multiply rapidly, and have a short life cycle. Since they are eukaryotes and more closely related to animals, their study is more applicable to human problems than is the study of bacteria. Fungi are used to study metabolite pathways, to study growth, development, and differentiation, for determining mechanisms of cell division and development, and for microbial assays of vitamins and amino acids. Fungi are also important genetic tools; the "one gene one enzyme" theory in *Neurospora* won George W. Beadle and Edward L. Tatum the Nobel Prize for Physiology or Medicine in 1958. The first eukaryotic genome to have its DNA sequenced was that of the bakers' and brewers' yeast, *Saccharomyces cerevisiae*.

III. HOW FUNGI GROW

A. Biology of Hyphae and Yeast Forms

A fungus is more than just the visible mushroom structure. In fact the mushroom, more properly called a fruiting body, is a very small portion of the individual life cycle and is mainly used for reproduction. The major portion of the life cycle, or the vegetative growth form, in the great majority of fungi consists of a system of threadlike, walled, more or less cylindrical hyphae (singular, hypha) making up what is called a mycelium (plural, mycelia) (Fig. 1). The Ascomycota and Basidiomycota have crosswalls called septa (singular, septum) separating compartments of the mycelium. An exceptional group is the yeasts, which consist of about 800 species that have a single-celled vegetative form. Note that yeast is a morphological term and has no taxonomic significance; yeasts and yeastlike forms can be found in all of the fungal phyla.

B. Exoenzymes and the Heterotrophic Lifestyle

Exoenzymes are the most important reason why fungi are so successful. Fungi excrete exoenzymes at the tips of the growing hyphae into their surrounding environment, where they play a major role in breaking down the substrate. Simpler molecules can then move into the hyphae by diffusion.

IV. REPRODUCTIVE BIODIVERSITY OF FUNGI

A. Fungal Life Cycles

The major events of any sexual life cycle are plasmogamy (cell and protoplast fusion), karyogamy (nuclear fusion), and meiosis (Fig. 2). In most other familiar types of organisms, such as plants and animals, plasmogamy and karyogamy occur in rapid succession and are usually referred to as the single event of fertilization. In the fungi, however, plasmogamy and karyogamy may be separated in time by several minutes, several hours, several days, several years, or even several centuries! Thus the dikaryon, the n+n stage, is a major component of the life cycle of fungi, especially in the Basidiomycota and Ascomycota. Nuclear cycles of all the members of the various phyla can be placed within this generalized nuclear cycle, differing mainly in the amount of time spent in each of the phases.

FIGURE 1 Hyphae with dark-stained nuclei.

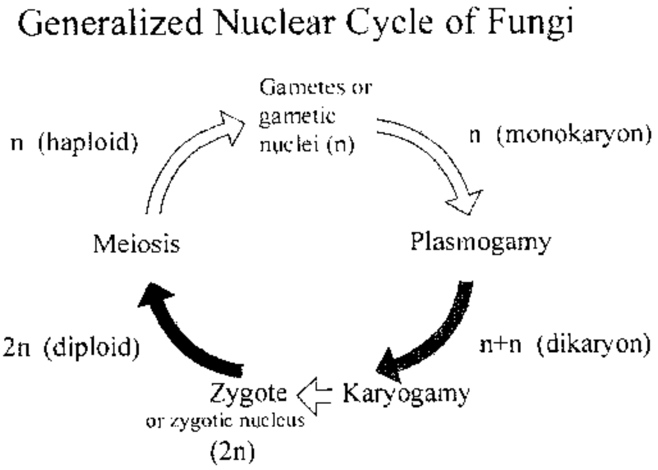


FIGURE 2 Generalized nuclear cycle for fungi.

Besides this sexual cycle, many fungi, commonly called molds, can also reproduce asexually (mitotically) in the absence of meiosis and karyogamy. Many of them produce specialized structures that bear the asexual spores. As in much of biology, there is some "competing" terminology here—the asexual state is also known as the anamorph or mitosporic state. Asexual reproduction can take place at any point in the life cycle (haploid, diploid, or dikaryon), depending on the species and conditions. The sexual state is also known as the teleomorph or meiosporic state.

B. Phyla of Fungi

Based primarily on variation in their sexual reproductive structures, the kingdom Fungi is usually divided into four major phyla.

- Chytridiomycota—sexual and asexual spores are motile, with posterior flagella.
- Zygomycota—sexual spores are thick-walled resting spores called zygospores, and asexual spores called sporangiospores (when present) are borne internally in structures called sporangia.

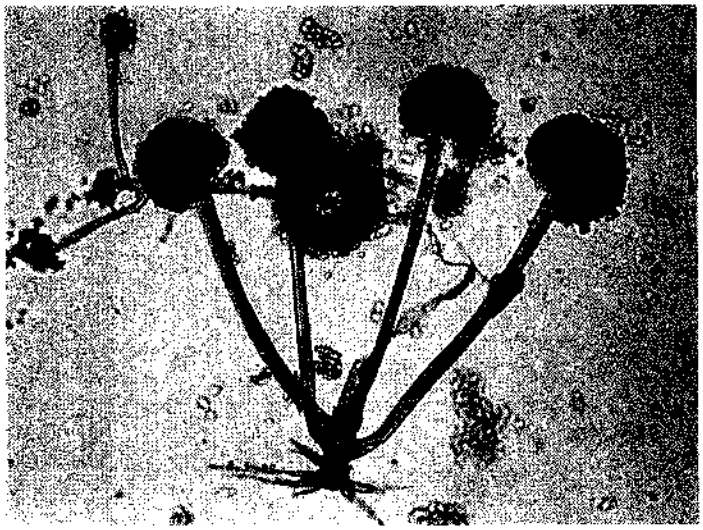


FIGURE 3 *Rhizopus* sporangia.

- Ascomycota—sexual spores (ascospores) are borne internally in a sac called an ascus, and asexual spores called conidia (when present) are borne externally on structures called conidiophores.
- Basidiomycota—sexual spores (basidiospores) are borne externally on a club-shaped structure called a basidium; most species do not have asexual spores, but when present they often take the form of conidia.

There is an additional group, the deuteromycetes, or "Fungi Imperfecti," for which there is no known sexual state. Its members have affinities to members of at least three of the phyla, especially the Ascomycota.

1. Chytridiomycota

The Chytridiomycota, commonly called the chytrids, are a group of mostly water-inhabiting organisms, although some are plant pathogens. In aquatic environments they mostly form scanty filaments with sporangia. Some examples of the Chytridiomycota are *Allomyces*, a water mold, *Synchytrium endobioticum*, a pathogen of potato, and *Neocallimastix*, a chytrid that lives symbiotically in the gut of herbivores, such as cattle. *Batrachochytrium dendrobatidis* and possibly other chytrids have been implicated as an infection associated with the worldwide decline in frog populations.

2. Zygomycota

Commonly called the bread molds, the Zygomycota are terrestrial fungi whose fruiting bodies are mostly microscopic in nature, although their asexually produced sporangia can reach greater than 5 cm tall in some species (Fig. 3). Under certain conditions they may sexually produce thick-walled resting spores called zygospores (Fig. 4). Some, such as *Rhizopus*, *Mucor*, and *Phycomyces*, can grow on a wide variety of substrates, and a few can act as human pathogens. Most importantly, members of one order, the Glomales, are responsible for forming mutualistic associations called endomycorrhizae with the roots of about 70% of the world's plants. Ectomycorrhizae (from Basidiomycota and Ascomycota) form with another 20% of plant species. See Section V,D,1 for further discussion of mycorrhizae.

FIGURE 4 *Syzygites* zygospore between suspensors.

FIGURE 5 *Morchella esculenta*, the morel, a prime edible mushroom. See also color insert, Volume 1.

3. Ascomycota

The Ascomycota bear their sexual spores (ascospores) in sacs called asci, which are usually cylindrical. Many

FIGURE 6 Morel asci containing ascospores.

members also form conidia as asexual spores. Familiar members of this phylum include the morels and other cup and saddle fungi, powdery mildews, the industrial yeast *Saccharomyces cerevisiae*, the incitant of chestnut blight (*Cryphonectria parasitica*,) the cause of Dutch elm disease (*Ophiostoma ulmi*), and a variety of other plant pathogens (Figs. 5 and 6).

4. Basidiomycota

The Basidiomycota bear their sexual spores externally on a usually club-shaped structure called a basidium, which is often borne on or in a fruiting body called a basidiocarp or basidiome (Fig. 7). This phylum includes the well-known mushrooms, both edible and poisonous, as well as puffballs, shelf fungi, jelly fungi, and coral fungi (Fig. 8). These species that produce fruiting bodies exhibit various methods of increasing their surface area, as discussed in Section IV,C.

FIGURE 7 Basidium showing two of the four basidiospores produced by meiosis and pinching off of the spores from the basidium.

The Basidiomycota also contain perhaps the most important plant pathogens, the rusts and the smuts. These fungi do not produce macroscopic fruiting bodies, but instead bear their spores on the stems, leaves, and flowers of host plants. However, remember that the mycelium is internal and "sucks" the nutrients out of the plant. Effects on the plant range from a reduced yield to death. Rusts in particular have very complicated life cycles, often requiring two unrelated host species to complete their growth stages.

5. "Deuteromycetes," the Fungi Imperfecti

The deuteromycetes, commonly called molds, are "second-class" fungi that have no known sexual state in their life cycle, and thus reproduce only by producing spores via mitosis. About 90% of these have affinities to the Ascomycota. Most food spoilage and fungal human diseases are caused by members of this group (Figs. 9 and 10). They are also known as the fungi imperfecti, because of their "imperfect" lack of sex. When the "perfect state" of one of these organisms is discovered, as happens every year, the fungus is more properly classified with the teleomorph name. Notice that this group is not classified as one of the phyla. It is just a loose assemblage of organisms that we are not sure where to place accurately in the taxonomic order.

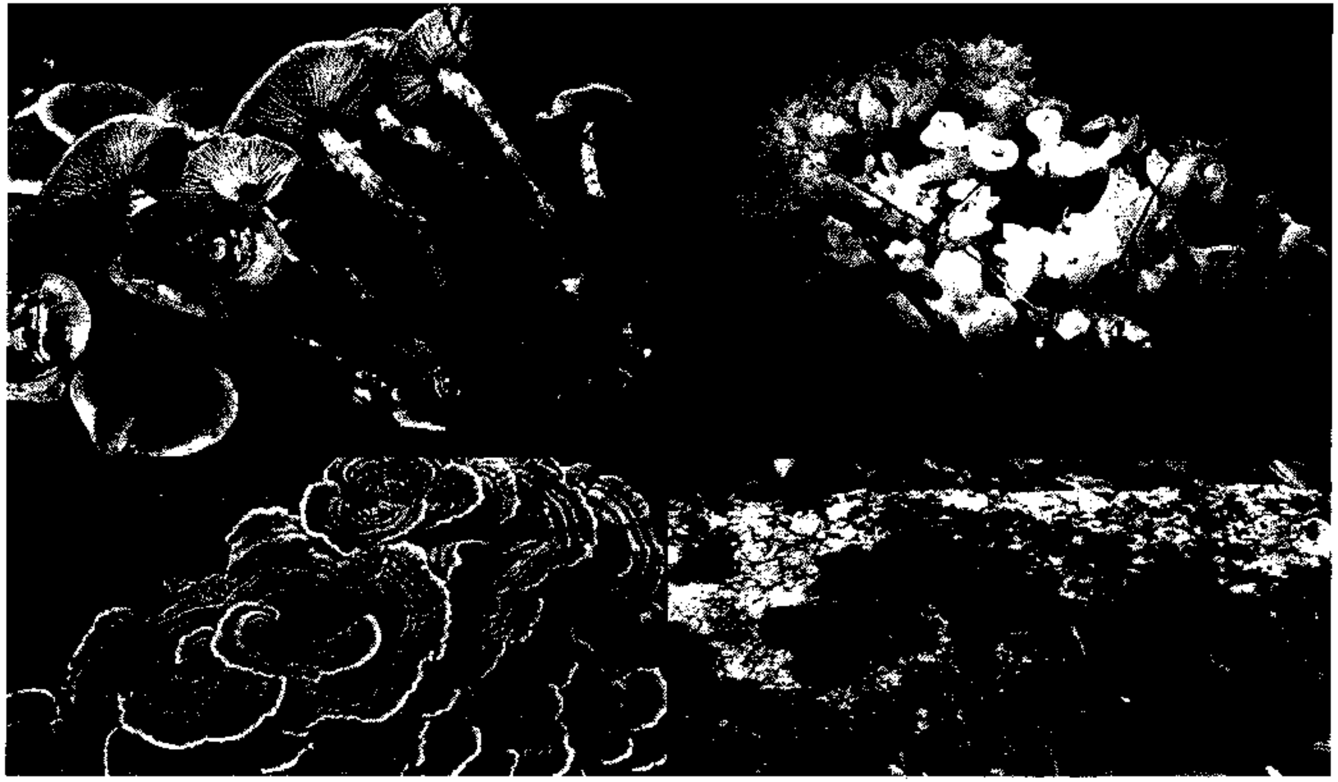


FIGURE 8 (A) *Armillaria nabsnona*, (B) *Tremella reticulata*, (C) *Trametes versicolor*, (D) *Pulcherricium caeruleum*. See also color insert, Volume 1.

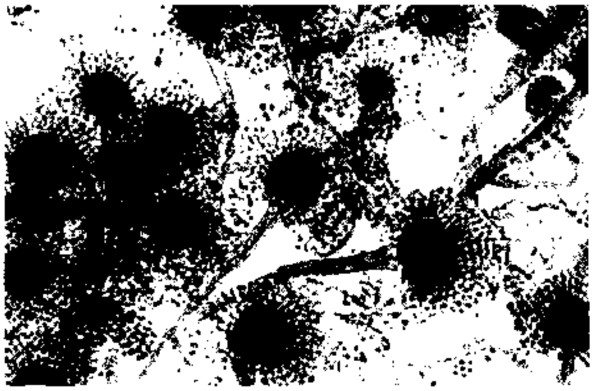


FIGURE 9 *Aspergillus* conidia. See also color insert, Volume 1.

6. Excluded Taxa: Former Members of the Kingdom Fungi

The organisms included here were once considered to be true fungi, but have been excluded because they do not exactly fit our modern classification scheme of the true fungi. They are now considered to belong in other kingdoms, such as Protista or Stramenopila.

a. Oomycota: Water Molds and Downy Mildews

The Oomycota have been excluded from the kingdom Fungi primarily because their cell walls are made of cellulose rather than chitin. They also have swimming zoospores, large nuclei, large egglike oospores, and various other unfungal-like features. They are closely related to the algae, simply lacking chloroplasts, and are now placed in the kingdom Stramenopila (or Protista according to some authors). Ecologically many of its members act like fungi, especially the plant pathogens. *Phytophthora infestans* causes a disease called late blight of potato, which was the cause of the Irish potato famine in the 1840s, in which more than a million Irish people perished and another million emigrated. The downy mildews *Peronospora parasitica* and *Plasmopara viticola* cause diseases of members of the cabbage family and of grapes, respectively. *Pythium* species cause damping-off disease of seedlings in agricultural practice. However, many other species are innocuous saprophytes that decompose debris in water. A few of these so-called water molds (some species of *Saprolegnia* and *Achlya*) are opportunistic fish parasites, especially in aquariums and fish hatcheries (Figs. 11 and 12). They are a particular problem in trout and salmon hatcheries.

b. Myxomycota: True Slime Molds

The Myxomycota are the true slime molds, also known as the plasmodial slime molds, and are considered members of the kingdom Protista. They exist in nature as a plasmodium—a blob of protoplasm without cell walls and only a cell membrane to keep everything in (Fig. 13). It is really nothing but a large amoeba and feeds much the same way, by engulfing its food (mostly bacteria) with pseudopodia, in a process called phagocytosis. So the slime mold ingests its food, then digests it. True fungi have a cell wall and digest their food with exoenzymes before ingesting it. When the plasmodium runs out of food it can form fruiting bodies. Most slime mold fruiting bodies are quite small, 1–4 mm in height, but

FIGURE 10 Jello mold, a favorite dish for picnics.

FIGURE 11 *Saprolegnia* oogonium containing oospores.

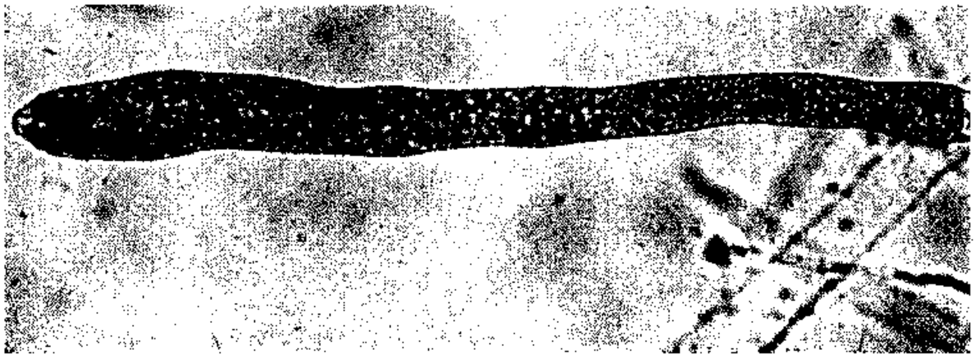


FIGURE 12 *Saprolegnia* sporangium containing small zoospores.

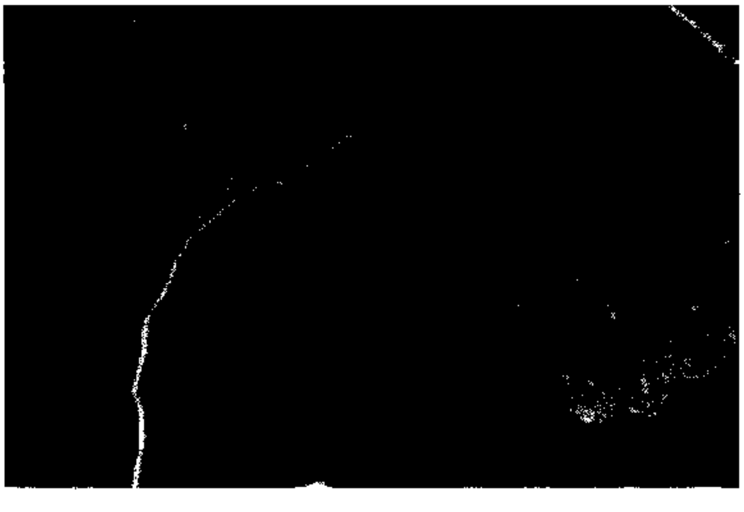


FIGURE 13 Plasmodium of a slime mold. See also color insert, Volume 1.

some can be up to 25 cm in diameter (Fig. 14). In such cases a large plasmodium may be seen crawling along the ground, inspiring science fiction movies such as "The Blob." These slime molds have been traditionally studied by mycologists because their small, delicate fruiting bodies tend to be fungal in appearance. Most slime mold fruiting bodies are quite beautiful.

c. Dictyosteliomycota: Cellular Slime Molds

The Dictyosteliomycota are the cellular slime molds or "social amoebae" and are among the most bizarre of microorganisms. These members of the Protista are free-living amoebae with no cell walls, indistinguishable from garden-variety amoebae until they begin to run out of food. At that point they signal to one another using cyclic AMP (a small nucleotide molecule) and begin to aggregate to form a "slug" or pseudoplasmodium. This slug, composed of hundreds of amoebae, acts as a single organism and can actually migrate along a light or temperature gradient. Eventually the slug stops migrating, rounds up, and forms a sorus, a kind of sporangium containing spores on a stalk. Not all of the cells become spores; some of them "sacrifice" themselves to become stalk cells to raise the spores up into the air for a better position for wind dispersal. So formerly free-living organisms act like a single organism for the good of the species. Very strange indeed!

d. Plasmodiophoromycota: Endoparasitic Slime Molds

The Plasmodiophoromycota are an odd group of endoparasites that live almost their entire life cycle inside a cell of another organism. They lack cell walls in the assimilative state and reproduce by means of swimming spores, the only part of their life cycle that does not occur inside a cell. *Plasmodiophora brassicae* causes club root in crucifers, and *Spongospora subterranea* causes powdery scab of potatoes. Scientists really do not know where to place these organisms taxonomically; they have been allied with the Oomycota, the Myxomycota, or various other protists, but in reality they are not closely related to any other known group of organisms. No one is even sure what kingdom to place them in!

C. Surface Area and Reproduction

Many fungi have very specialized habitats. For example, the basidiomycete *Suillus americanus* can grow only in association with the eastern white pine, *Pinus strobus*. To infect a new white pine, the immobile mycelium of the fungus must produce spores to move to a new host. These spores are carried by the wind, not by some specific insect or other animal vector. Thus the fungus must produce enormous numbers of spores so that a few of them will land on the "correct" substrate. The underside of the *Suillus* mushroom has small pores that are lined with microscopic spores. This increases the surface area for bearing spores by more than 100 times. More commonly, mushrooms have gills underneath the cap to increase surface area. Other fungi increase their surface area by forming upright coral-like branches, whereas others form downward-pointing teeth or spines

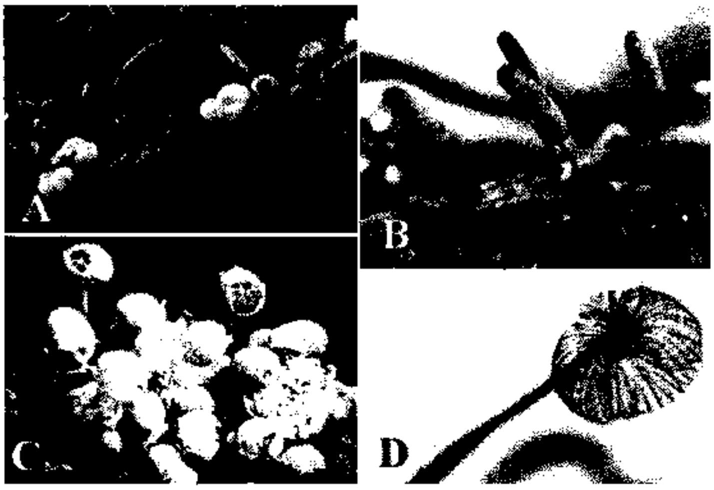


FIGURE 14 (A) *Lycogala epidendrum*, (B) *Comatricha typhoides*, (C) *Badhamia utricularia*, (D) *Dictydium cancellatum*.

FIGURE 15 Different ways of increasing the surface area for bearing spores. Clockwise from upper left: pores, upright branches (coral), teeth, and gills.

(Fig. 15). Others (not shown) have blunt ridges and some have folds, and some increase the time of sporulation rather than the spore-bearing physical space. Analogous surface area modifications are made by members of other phyla of fungi.

V. FUNGI IN THE ENVIRONMENT

Fungi occupy many different niches in the environment, although any one species usually occupies only a single niche. Many niches have not yet been explored for fungi, which is the major reason why mycologists believe there are high number of fungi yet to be discovered. Fungi can be divided into groups based on their nutritional status and the nature of their relationship with their host. Saprophytes use nonliving organic material and are important scavengers in ecosystems. Along with bacteria, fungi are important in recycling carbon, nitrogen, and essential mineral nutrients. Parasites use organic material from living organisms, harming them in some way. Their hosts range from single-celled diatoms to other fungi, plants, animals, and humans. Fungi are the major parasites (pathogens) of plants. Mutualists are fungi that have a reciprocally beneficial relationship with other living organisms, in which both organisms benefit. The two main types of mutualistic associations involving fungi are mycorrhizae, which are associations of fungi with plants roots, and lichens, which are associations of fungi with algae or cyanobacteria There are also a few commensal fungi that use other organisms as merely a place to live; these fungi derive no nutrition from their host.

A. Fungi as Saprophytes

Along with bacteria, fungi are the major decomposers and recyclers in the environment. For every sort of dead material present, there is usually at least one fungus that can degrade that material. A few exceptions include some pesticides and some types of plastics; no fungi have yet developed exoenzymes capable of digesting these synthetic materials. Fungi are important in breaking down carbon- and nitrogen-containing compounds into components that they and other organisms can use. Fungi are especially important in the breakdown of the wood components cellulose and lignin (discussed in Section VI,F).

B. Fungi as Plant Parasites

About 90% of plant diseases are caused by fungi, resulting in billions of dollars in crop and forest losses each year. A parasite is referred to as a pathogen if it elicits a recognizable and persistent disease. The most severe pathogens are those that are imported accidentally from other continents and that cause severe problems with the native populations and cultivated plants. The severity of any plant disease is determined by three factors, known as the disease triangle, consisting of the interaction of the host (conditions favoring susceptibility), pathogens (conditions affecting virulence), and the environment (conditions favoring disease)(Fig. 16). All three factors must work in conjunction to produce a disease and determine its harshness.

1. Fungi in Crop Pathology

The science of plant pathology deals mostly with organisms, especially fungi, that cause plant disease. Plant diseases are generally classified on the basis of what kinds of symptoms occur in which part(s) of the plant (Table II). Some important fungal pathogens of crop plants include *Puccinia graminis* (black stem rust of wheat), *Erysiphe graminis* (powdery mildew of grasses), *Claviceps purpurea* (ergot), and *Ustilago maydis* (corn smut).

2. Fungi in Forest Pathology

There have been a number of fungal diseases of forest trees that have caused great problems in North America (Table III). Forest pathogens often need a longer time to kill their larger, perennial hosts, sometimes living inside the tree for several decades before the host finally succumbs.

Cryphonectria parasitica, *Ophiostoma ulmi*, and *Discula destructiva* are introduced pathogens that have dev-

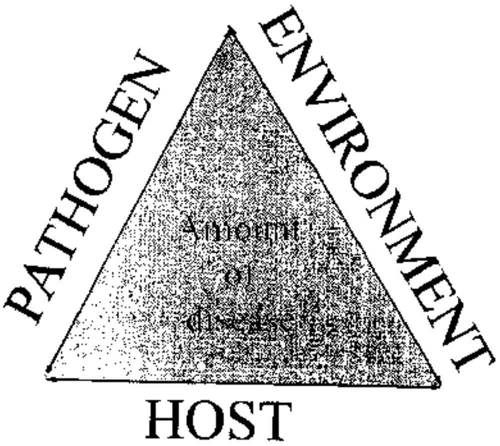


FIGURE 16 The disease triangle, showing the three factors that determine the disease severity.

TABLE II

Types of Plant Diseases

Type of disease	Symptoms
Blight	Extensive necrosis and rapid death
Dieback	Dead branches protruding from apparently healthy tissue
Lesion	Localized cell death
Canker	Lesion on a woody stem
Rot	Extensive decay of tissue
Vascular wilt	Blockage of the vascular transport system, resulting in starvation or disiccation
Gall	Localized tumorlike growth
Stunting	Reduction in overall growth size
Rusts	Rust-colored lesions on various parts of the plant
Smuts	Drastic malformations, usually of reproductive parts

astated their hosts in North America. In their native Asia, these pathogens coevolved with their hosts and caused them only minor annoyance. However, when they were accidentally introduced into North America, either on live plants or on logs for sawing, the fungus escaped and subsequently devastated the North American tree species because there was not enough time for the host to develop resistance. *Cronartium ribicola*, *Ceratocystis fagacearum*, and *Armillaria* species are native American fungi, but they cause severe diseases nonetheless. A good pathogen does not kill its host right away, but keeps it alive as long as possible to continue deriving nutrients from it.

C. Fungi as Animal Parasites

1. Fungi Affect Humans

Fungal effects on humans are classified by whether they can grow on the body (mycoses) or whether they cause problems by their ingestion. Mycetismus refers to the eating of poisonous mushrooms, and mycotoxicosis refers to the ingestion of toxins produced by the fungus, not involving eating the fungus itself. Mycoallergies may be a separate category.

a. Mycosis: Fungi Growing Directly on Human Tissues

There are a number of fungi that can grow directly on human tissue. Most of them require some debilitation of host defenses, especially the immune system. Fungal diseases have gained in importance over the last couple of decades because of increases in the numbers of such patients, including those with AIDS, people on corticosteroid therapy or immunosuppressive drugs (such as the transplant drug cyclosporin), diabetics, and people undergoing chemotherapy treatment for cancers.

About 175 human pathogens are recognized among the approximately 70,000 known species of fungi. Around 20 are regularly isolated from cutaneous infections (dermatophytes and yeasts), a dozen are associated with severe subcutaneous localized disease, and

TABLE III

Some Important Forest Pathogens

Pathogen	Disease
Cryphonectria parasitica	Chestnut blight
Ophinostoma ulmi	Dutch elm disease
Discula destructiva	Dogwood anthracnose
Cronaritum ribicola	White pine blister rust
Ceratocystis fagacearum	Oak wilt
Armillaria spp., especially *A. mellea* and *A. ostoyae*	*Armillaria* root rot

about 20 may cause systemic infections (Table IV). In addition, there are many opportunistic pathogens that cause disease in debilitated or immunosuppressed patients.

Probably the most common fungal infection in North America is yeast infection, caused by *Candida albicans*. This yeast can grow on outside portions of the skin or in any area of the body that is moist and warm. It is especially prevalent in the genital area of women. Oddly enough *Candida* is part of the normal flora of the body and can be isolated from almost everyone at any time. Only when conditions get out of balance does the yeast flourish and cause disease. However, the systemic true pathogens are the most feared, since they do not require a debilitation of the host defenses to become pathogenic. These dimorphic fungi get around host defenses by changing their form from a mycelium to a yeast upon change in temperature, thus evading even the healthiest of immune systems. Fortunately they are all geographically restricted. (see also Rippon, 1988.)

b. Mycetismus: Eating Poisonous Mushrooms

Of the 70,000 species of fungi, there are about 250 species of delicious edible mushrooms and about 250 species that will kill you. Most of the other species are innocuous or unpleasant tasting. Some people call poisonous mushrooms "toadstools," a word probably derived from the old German *Todstuhle* for death chair. But how can you tell the difference? There is no easy way to tell if a mushroom is poisonous. Old wives' or old husbands' tales about silver spoons or whether it grows on wood or whether animals eat the mushrooms do not work! The only way to be sure if a mushroom is edible is to identify it to species by consulting books or knowledgeable persons. The old saying is "there are old mushroom hunters, and there are bold mushrooms hunters, but there are no old, bold mushroom hunters." The best way to learn about edible mushrooms is to join a local mushroom club. Do an internet web search for the North American Mycological Association for a list of these clubs. (NOTE: The pictures and descriptions in this article are NOT adequate to identify mushrooms for eating. We have a report of a woman who nearly died from eating wild mushrooms that she identified only from an encyclopedia drawing. Don't let this be you! Be absolutely sure of your identifications!)

As shown in Table V, unrelated species in various genera may contain the same toxin, and one or more toxin groups may be found in the same genus. To complicate things further, many genera contain both poisonous and nonpoisonous species, although several genera are more or less homogeneous with respect to poisoning ability. Most species of fungi have not yet been analyzed for toxins. Evidence suggests that mushroom toxins have not evolved for protection. For example, what good does it do a mushroom such as *Amanita virosa*, known as the death angel, if an animal eats the mushroom and then dies two days later? The toxins

TABLE IV

Mycoses Can Be Characterized by Their Location on the Body

Type of mycosis	Location	Some diseases	Some fungi causing disease
Superficial mycoses	Infections of the hair shaft or dead outer layer of skin (stratum corneum)	Pityriasis versicolor, tinea nigra palmaris, piedra	*Malassezia furfur*, *Trichosporon beigelii*, *Cladosporium*, *Piedraia hortae*
Cutaneous mycoses	Dermatophytes—infections of skin, hair, and nails	Ringworm, jock itch, athlete's foot, tinea corporis, tinea capitis, and many others	*Microsporum*, *Trichophyton*, and *Epidermophyton* spp., occasionally a *Candida* sp.
Subcutaneous mycoses	Chronic localized infections of the skin and subcutaneous tissues	Sporotrichosis, chromoblastomycosis, phaeohyphomycosis	*Sporothrix schenckii*, *Phialophora*, *Cladosporium*
Systemic mycoses (deep mycoses), true human pathogens	Fungal infections of the body caused by dimorphic fungal pathogens, usually entering through the lungs	Histoplasmosis Blastomycosis Coccidiodomycosis Paracoccidioidomycosis	*Histoplasma capsulatum* *Blastomyces dermatitidis* *Coccidiodes immitis* *Paracoccidioides brasiliensis*
Systemic mycoses, opportunistic pathogens	Fungal infections of the body caused by common fungi becoming pathogenic owing to patient debilitation	Aspergillosis Cryptococcosis Candidiasis Zygomycosis *Pheumocystis* pneumonia Penicillosis	*Aspergillus* spp. *Cryptococcus neoformans* *Candida* spp. *Mucor*, *Rhizopus* *Pneumocystis carinii* *Penicillium marneffei*

TABLE V

Seven Major Classes of Mushroom Toxins

Class of toxin	Fungi where found	Main symptoms	Mode of action
Cyclopeptides (amatoxins and phallotoxins)	*Amanita virosa* (Death angel), *A. phalloides* destroying angel), *A. verna*, *A. ocreata*, *Galerina autumnalis*	Violent cramps, diarrhea, nausea, jaundice, coma, death; it normally takes about 12–36 hours for symptoms to appear	Attacks RNA polymerase in the liver, eventually destroys liver
Monomethyl hydrazine (gyromitrin)	*Gyromitra* species (false morels)	Nausea, vomiting, diarrhea, cramps, jaundice, convulsions, coma, death; cancers produced in mice	Similar in structure to solid rocket fuel, destroys red blood cells, attacks central nervous system
Coprine (tippler's bane)	*Coprinus atramentarius*	Hot, flushed, metallic taste, palpitation, vertigo, vomiting headache like a bad hangover	Blocks alcohol dehydrogenase so alcohol cannot be completely metabolized; similar in structure and function to Antabuse
Muscarine	*Inocybe* spp., *Clitocybe dealbata* (the "sweater")	Induces PSL (perspiration, salivation, lacrimation) symptoms, cramps, blurred vision; contraction of pupils, hypotension	Anticholinergic—antagonist to action of parasympathetic nerve fibers
Ibotenic acid, muscimol	*Amanita muscaria* and *A. pantherina*	Rusults in "expanded perception," talking to God, macropsia (perceiving objects as enlarged), rapid heartbeat, dry mouth	Hallucinogenic, psychoactive, action on nervous system, acts as a neuropeptide receptor
Psilocybin, psilocin, psychedelic mushrooms	Many *Psilocybe* spp., some *Panaeolus* spp.	Trancelike state induced, hallucinations	Serotonin (neurotransmitter) analog
Gastrointestinal irritants	Many different kinds, found in hundreds of species	Nausea, vomiting, no effect on other organs, most just cause discomfort or distress to varying degrees	Various, depending on the species of mushroom

appear to be merely waste products, usually from nitrogen metabolism, that happen to be poisonous to animals or humans.

c. Mycotoxicosis: Eating Toxins Produced by Fungi, not Necessarily from Eating the Fungi Themselves

Mycotoxins are usually produced by molds growing on foods. As the molds grow, they metabolize the food product and excrete their waste products back into the substrate. Some of these waste products happen to be highly toxic, and when a person eats the toxin (even after cutting the mold off), there is usually some harm to the person. The common contaminants *Aspergillus flavus* and *A. parasiticus* can produce aflatoxins, especially in peanuts. If you have ever eaten a whole peanut that tastes so bitter that you spat it out, it was probably contaminated with aflatoxin. This compound is highly carcinogenic at about 10–20 parts per billion, and kills very rapidly at higher concentrations. There are legal limits for the amount of aflatoxin allowed in peanut butter sold in the United States.

2. Fungi Affect Other Animals

Nonhuman animals are affected by fungi in much the same way that humans are, with some differences in certain species. For example, dogs are susceptible to cyclopeptide poisoning while cats apparently are not. There are numerous other examples of this discrepancy—so do not pick and eat a mushroom just because you see an animal eat it. Another important point to consider is that you do not know what happened to that squirrel after it ate that mushroom.

Animals are also affected by many of the same fungal diseases as are humans. For example, dogs are particularly susceptible to blastomycosis and are often used as a warning sentinel, like a canary in a mine, for alerting humans to possible risk.

D. Fungi as Mutualists with Other Organisms

Rather than being harmful, some fungi benefit their host in some way while receiving nutrients from them.

The two most common mutualistic associations are mycorrhizae, an association between a fungus and the roots of a plant, and lichens, an association between a fungus and either an alga or a cyanobacterium or both.

1. Mycorrhizae

According to Harley and Smith (1983), a mycorrhiza is defined as "an association between a fungus and a host plant in which destructive disintegration of the host does not occur and which is a prevalent and usual condition of the host plant in natural habitats and as such is very common and widespread." *Myco*, of course, means "fungus," and *rhiza* is "root," so mycorrhiza literally means "fungus root." More than 90% of plants in nature have a mycorrhizal symbiont. The only groups of plants that regularly lack mycorrhizae are some crucifers, sedges, and some legumes.

There are several types of mycorrhizae. In ectomycorrhizae, the fungus, usually a Basidiomycota or sometimes an Ascomycota, forms a sheath outside the root (Fig. 17). Exchange of nutrients takes place in transfer cells called a Hartig net, which penetrates between the cells of the root cortex, but does not penetrate the cells themselves. In endomycorrhizae, also called vesicular-arbuscular mycorrhizae (VAM) or sometimes simply arbuscular mycorrhizae, no sheath is formed. This fungus is always a member of the Zygomycota. Nutrient exchange takes place in highly branched hyphae called arbuscules, which penetrate into the cortical cells, but do not penetrate the cell membrane. Orchid mycorrhizae and ericoid mycorrhizae are special types that are found with plants in the Orchidaceae and the Ericaceae, respectively, and they differ significantly in their structure and life strategies.

In all types of mycorrhizae, hyphae extend from the root into the surrounding soil, greatly increasing the surface area for absorption of nutrients, particularly phosphate, nitrogen, and potassium. In return for shunting some of these nutrients into the plant, the fungus receives some sugars from plant photosynthesis. Thus both organisms benefit. Mycorrhizal fungi are abundantly represented in fossils from the Devonian and later periods and apparently coevolved with their hosts. It has been hypothesized that these fungi were necessary for the movement of water plants onto land.

As an interesting sidelight, there are also several hundred species of nonphotosynthetic plants (such as *Monotropa uniflora*, the Indian pipe) that get their energy as parasites of fungi that are mycorrhizal with photosynthetic plants. Radioactive carbon has been used to trace nutrient flow from the host plant through the mycorrhizal fungus and into the achlorophyllous plant.

2. Lichens

A lichen is a dual organism that consists of a mutualistic relationship between a fungus (the mycobiont) and an alga or cyanobacterium (the photobiont). Usually neither can survive on its own. Most of the fungi involved are Ascomycota, though a few are Basidiomycota. There are about 16,000 species of lichens, many of which can grow in very inhospitable environments—on rocks, sides and branches of trees, and gravestones, from the tropics to deserts to the Arctic (Fig. 18). Lichens are

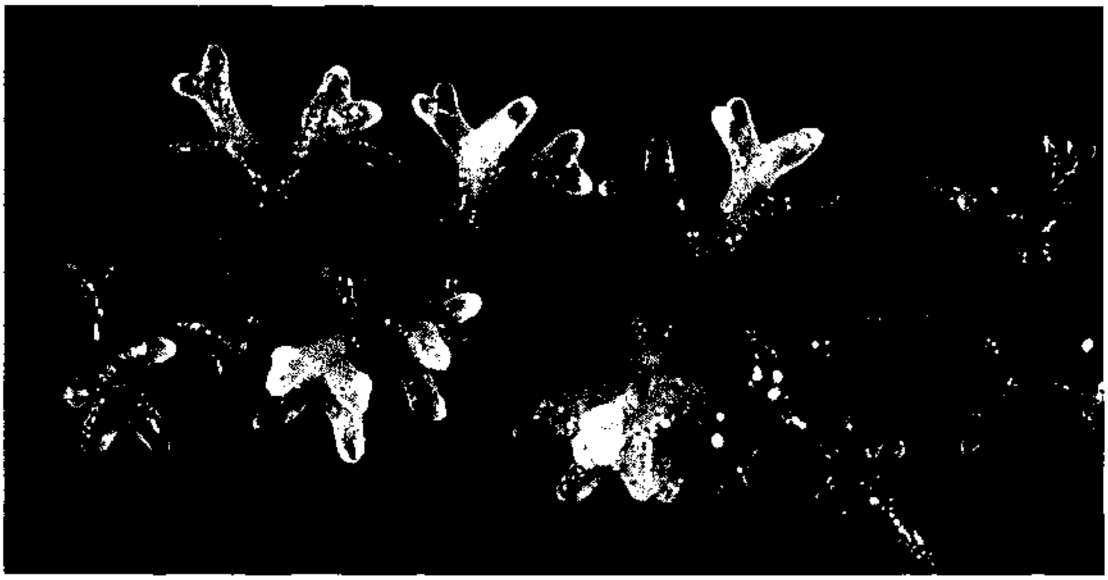


FIGURE 17 Ectomycorrhizae on pine roots.

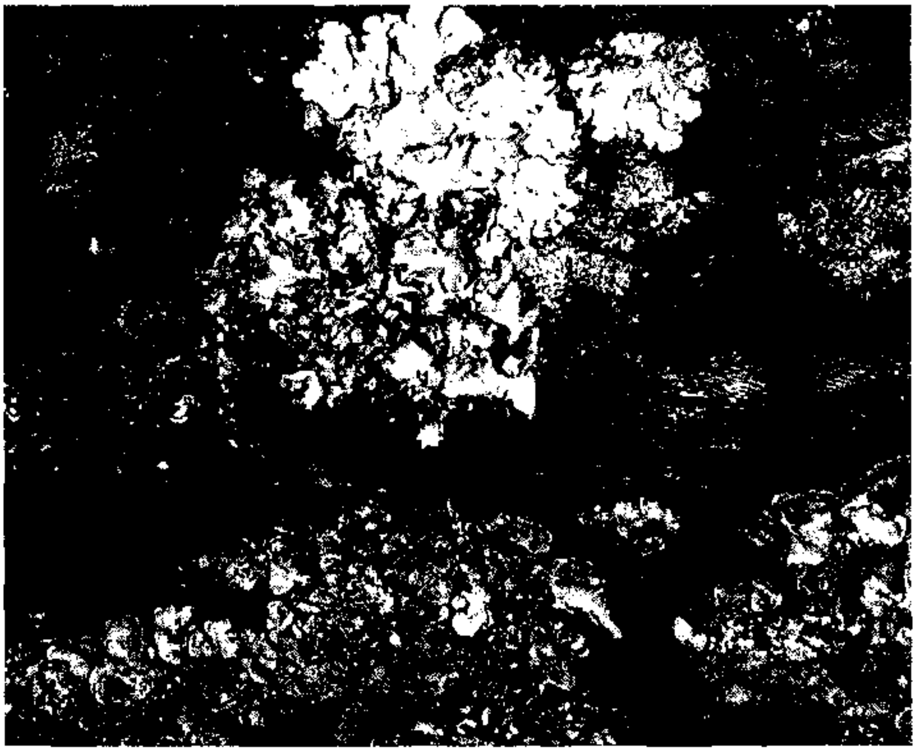


FIGURE 18 Some foliose lichens.

very sensitive to air pollution, especially sulfur and nitrogen, and so they are natural indicators of air quality. However, even under optimal conditions, lichens grow extremely slowly, usually 1–2 mm per year. The main ecological importance of lichens is their capacity to break down rocks into soil. They are an important food source for caribou and reindeer on the tundra. One of the first indications to the outside world that there had been a nuclear accident at Chernobyl, then in the Soviet Union, was the accumulation of radioactivity first in the lichens in Scandinavia, then in the milk given by the reindeer. There is significant evidence that the manna referred to in the Bible was actually a lichen, *Lecanora esculenta*, a loosely attached lichen that was blown down from the mountains into the lowlands. Some lichens have been used as natural dyes, such as tweed. In less enlightened times, a bright yellow lichen called *Letharia vulpina* was used by "pioneers" as a wolf poison.

VI. INDUSTRIAL USES FOR FUNGI

A. Mushroom Cultivation

The most widely available mushroom produced in the United States is the white button mushroom, *Agaricus bisporus* (Table VI). Commonly used on pizza and at salad bars, this mushroom is a secondary decomposer that is grown on composted cow or horse manure. Some brown forms of *A. bisporus* are currently being cultivated; brown buttons are sold as crimini and opened mushrooms are sold as portabella mushroom.

There are a number of steps in the commercial production of *Agaricus bisporus:*

- Manure is placed in large concrete "runways."
- Composting (breakdown of the substrate into simpler components) occurs with the bacteria and fungi naturally present in the manure.
- Large machines turn the compost weekly, otherwise the center of the pile gets too hot because of metabolic heat and kills the composting bacteria and fungi. This also ensures even composting. The different species present naturally shift as composting progresses.
- After a couple weeks, the odorless compost is ready and is placed into large trays, 6 ft. × 6 ft. × 2 ft.
- Mycelium of *Agaricus* is inoculated into the composted substrate and allowed to grow for a few weeks.
- When the substrate is colonized, a sterile layer of

TABLE VI

Some Commonly Cultivated Edible Mushrooms

Common name	Scientific name
White button pizza mushroom, crimini, portabella	*Agaricus bisporus* (a.k.a. *A. brunnescens*)
Shiitake, shiang-gu	*Lentinula edodes*
Oyster mushrooms	*Pleurotus* spp.
Enoki, velvet stem, winter mushroom	*Flammulina velutipes*
Maitake, hen-of-the-woods, sheepshead	*Grifola frondosa*
Cloud ear, wood ear, black mushroom	*Auricularia auricula*, available only dried
Chinese paddy straw mushroom	*Volvariella volvacea*, available only canned
Pom-pon or Lion's Mane	*Hericium erinaceus*
Reishi	*Ganoderma lucidum*
Wine cap	*Stropharia rugoso-annulata*
Morels	*Morchella* spp.

nutrient-poor casing soil is placed over the top of the substrate.

A few weeks later, the mycelia send up rhizomorphs (hyphal aggregations) through the casing layer and form mushrooms. Contrary to popular belief, most mushrooms require light to initiate fruiting body formation. Light is often used as a signal to the mycelium that it is outside the substrate and that fruiting bodies can be formed. Some fungi form fruiting bodies in response to reaching outside air, where the concentration of carbon dioxide is lower. A few fungi fruit only when they run out of available nutrients. *Agaricus bisporus* has been bred so that it requires no light for fruiting.

There are a number of other specialty mushrooms being grown throughout the world that are just becoming available in North American supermarkets. Almost all the others are primary decomposers of wood or other cellulose-containing substrates. These have gained in popularity in the past 10 to 20 years as consumers discover that these specialty mushrooms have great flavor and other interesting qualities. In Japan, China, and Korea, very little *Agaricus* is being grown; growers concentrate on mushrooms with more robust flavor and interesting texture, such as shiitake (Fig. 19). Most of these specialty mushrooms are now grown on artificial sawdust logs by the following method:

- Place sawdust or wood chips with supplementary bran and millet and the appropriate amount of water into clear polypropylene bags. Autoclave or sterilize the filled bags, then allow to cool.
- Inoculate spawn (usually grain or sawdust with mycelium of the fungus growing on it) into medium. Mix thoroughly by hand or mechanically.
- Place bags in growth room. Allow spawn to grow rapidly through the substrate. Fungus colonizes bag in 30–60 days; sawdust is easier than solid wood to colonize because of increased surface area-to-volume ratio, abundance of air spaces, and uniform distribution of nutrients.
- During this period the loose medium is joined together into a coherent synthetic log.
- Depending on the species, at this time the plastic is removed, and the synthetic log can be handled like a natural log. Fruiting usually occurs within 90–120 days after inoculation.

B. Antibiotics and Other Drugs

Penicillin is the first antibiotic that was discovered to fight the bacteria that cause human disease. It is naturally produced by *Penicillium chrysogenum* and related species as way of killing bacterial competitors in their environment. Cephalosporins are another class of antibiotics produced by *Acremonium* and related species. For a fungal infection of the fingernails and toenails, one prescribed drug is griseofulvin, produced by *Penicillium griseofulvum*. Ergotamine, produced by *Claviceps purpurea*, is used to facilitate the delivery of babies and can also be used to relieve migraine headaches. Another chemical found in *Claviceps* is a precursor to the hallucinogen LSD (lysergic acid diethylamide) that has the same effects as that illegal drug. The steroids in birth control pills are produced industrially by the fungus *Rhizopus nigricans*, as are the steroids cortisone and

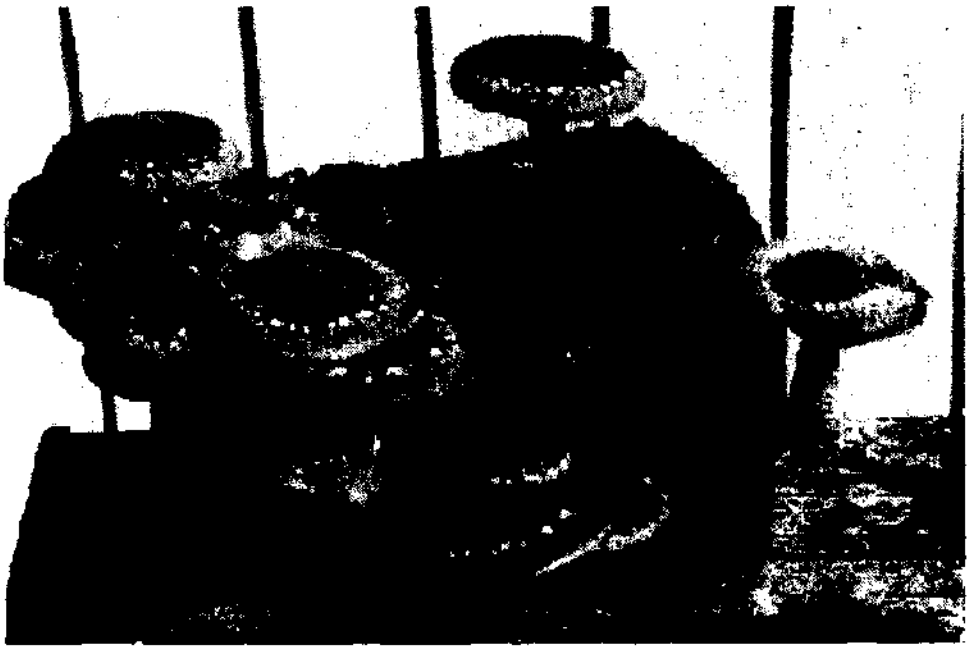


FIGURE 19 Shiitake mushrooms fruiting on an artificial log.

prednisone. People who have had an organ transplant usually take the antirejection drug cyclosporin, which is produced by the fungus *Tolypocladium inflatum*. The pharmaceutical industry is constantly searching for new antibiotics and drugs to counteract microbes that become resistant to frequently used medications.

C. Wine and Beer Making

Yeasts, especially *Saccharomyces cerevisiae*, have played an important role in the development of human culture. Since brewing is an ancient art going back for millennia, yeasts have been our allies for many years. Virtually anything that contains simple sugars is fermentable by yeasts. In wine making, the grapes are pressed and yeast is added. Fermentation breaks the sugars down into carbon dioxide and ethyl alcohol. Since barley is the main carbohydrate source in beer making, there is an added step before fermentation, that of allowing the barley to germinate to produce sugar from its starch. Various other ingredients, such as hops, oatmeal, and flavorings, are added to make the wide variety of beers that are available. The aging process in alcoholic beverages also adds to their distinct flavors.

D. Bread Baking

Bread making also uses *Saccharomyces cerevisiae* to ferment sugars into carbon dioxide and ethyl alcohol. However, bakers are only interested in the carbon dioxide, which makes the bread rise. The alcohol evaporates rapidly on baking.

E. Fermentation Products for Food Use

Fungi are often used in the large-scale fermentation of liquid or solid substrates. Fermentation vats can sometimes be several hundred thousand liters in size. For example, citric acid in cola drinks is produced by large-scale vat fermentation of the deuteromycete *Aspergillus niger*. Yeasts are sometimes grown in large fermenters and used directly as food supplements. Vitamin B_2 (riboflavin) in enriched flour is produced by the ascomycete *Ashbya gossypii*. Cocoa beans are processed by a "fermentation" (sensu food scientists) of *Candida krusei* and *Geotrichum*. Authentic soy sauce is fermented in a three-step process with the fungi *Aspergillus oryzae* and *Zygosaccharomyces rouxii*, as well as the bacterium *Pediococcus halophilus*. Tempeh, a soybean product popular in Indonesia, is partially fermented with a species of *Rhizopus*. Many good cheeses, such as blue cheese, camembert, and brie, are ripened through the action of fungi to obtain their distinctive flavors. Blue cheeses such as Roquefort, Gorgonzola, and Stilton are ripened by *Penicillium roquefortii*—the blue color is caused by sporulation of the fungus! The white crust on the outside of brie and camembert is the mycelium of *Penicillium camembertii*.

F. Biopulping and Bioremediation

Several important industrial process, which have the potential to be great boons to ecosystem health, are in the pilot stage: biopulping and bioremediation. The lignin-degrading enzyme system of *Phanerochaete chrysosporium* is special for these two uses. One of the biggest energy and pollution expenditures in paper making comes from removal of the brown lignin from wood so that the white cellulose is all that is left to make paper. What if paper companies could use the enzymes of a fungus to remove the lignin? This could result in a savings in both energy and time and avoid the polluting wastes that are commonly dumped out of the mills. This process is known as biopulping. There are several products in the pilot stage, but no large-scale biopulping is yet being done.

To understand this system, you must know that wood consists primarily of cellulose, which is white, and lignin, which is brown. *Phanerochaete* species cause a white rot of wood. That is, the fungus decays the lignin and leaves the cellulose behind. There are also fungi that cause a brown rot, digesting the cellulose and leaving the lignin behind. Many kinds of fungi cause a white rot, but *P. chrysosporium* has several features that might make it very useful. First of all, unlike

FIGURE 20 The author with *Bridgeoporus nobilissimus* on a large host tree of *Abies procera*.

some white rotters, it leaves the cellulose of the wood virtually untouched. Second, it has a high optimum temperature (about 40°C), which means it can grow on wood chips in compost piles, which attain a high temperature. These characteristics point to some possible roles for this fungus in biotechnology applications.

Some of the lignin-degrading enzymes of *P. chrysosporium* will also degrade toxic wastes, such as PCBs (polychlorinated biphenyls), PCPs (phencyclidines), and TNT (trinitrotoluene). The structure of these chemicals is similar to that of lignin, and the ligninase enzymes will work on them. The fungus performs well on the laboratory bench, but as with many industrial bioprocesses, there are difficulties in scaling up the process. Nonetheless, this procedure has the potential to clean up some industrial and toxic waste sites.

VII. THREATENED OR ENDANGERED FUNGI

Like the proverbial canary in the coal mine, fungi may be the first indicators of things going wrong in ecosystems. Remember that fungi are the thread that tie the whole food web together, since they are the primary decomposers and aid most plants as mycorrhizae in the absorption of minerals and water. Thus if something goes wrong with the fungi—if they disappear—there may be dire consequences for any plants and animals that are dependent on them. Lichens have already been proven to be accurate indicators of air quality, in both their quantity and diversity.

Several European countries maintain "red lists" of threatened or endangered fungi. One fungus that may be a candidate for the endangered species list in the United States is *Bridgeoporus nobilissimus*, a polypore fungus with a very large, perennial fruiting body (Fig. 20). For a long time this fungus was in the *Guinness Book of World Records* as the largest known fruiting body of a fungus, at over 160 kg (300 pounds)! There are just six known sites in Washington and Oregon at which *B. nobilissimus* is now known to occur. It is considered by many to be a rare and probably endangered fungus. The main reason for this designation is that it is restricted to very large specimens of noble fir (*Abies procera*) and occasionally Pacific silver fir (*Abies amabilis*) with a diameter at breast height (dbh) of 1–2 meters. Trees of this diameter are not very common. This fungus is considered endangered because its habitat is endangered. Unfortunately, it is not clear whether the U.S. Endangered Species Act applies to fungi. In this zoocentric world, most people are more interested in the "charismatic megafauna" than in some lowly fungus. However, the fungi have many important roles to play in the ecosystem and should not be ignored. No ecosystem could exist for long without fungi!

Two states have thus far recognized the importance of fungi by naming a state mushroom. Minnesota was the first, naming the morel (*Morchella* species) as their state mushroom. Some ten years later, Oregon declared the Pacific golden chanterelle (*Cantharellus formosus*) as their state mushroom. Surprisingly, some forests yield more income annually from wild mushroom harvesting than from lumber harvesting. Moreover, intact forests continue to produce edible mushrooms yearly, whereas lumber harvesting can occur only once every 50 years or so. Mushrooms for human food is only one of the many contributions that fungi make to our lives, for their fundamental decomposer role supports almost all ecosystems and so helps to provide the essential ecological services that we take for granted, as well as the recreational opportunities of enjoying nature. Fungi have contributed a great deal to our standard of living by making most of the living world possible.

See Also the Following Articles

EUKARYOTES, ORIGIN OF • PARASITISM

Bibliography

Alexopoulos, C. J., C. Mims, and M. Blackwell. (1996). *Introductory Mycology*. John Wiley & Sons, New York.

Barron, G. (1999). *Mushrooms of Ontario and Eastern Canada* (also published as Mushrooms of Northeast North America). Lone Pine Publishing, Edmonton, Alberta, Canada.

Benjamin, D. R. (1995). *Mushrooms: Poisons and Panaceas*. W. H. Freeman & Company, New York.

Bessette, A., A. R. Bessette, and D. Fischer. (1997). *Mushrooms of Northeastern North America*. Syracuse University Press, Syracuse, New York.

Castlebury, L. A., and L. L. Domier. (1998). Small subunit ribosomal RNA gene phylogeny of *Plasmodiophora brassicae*. *Mycologia* 90, 102–107.

Farr, D. F., G. F. Bills, G. P. Chamuris, and A. Y. Rossman. (1989). *Fungi on Plants and Plant Products in the United States*. APS Press, St. Paul, Minnesota.

Gilbertson, R., and L. Ryvarden. (1986, 1987). *North American Polypores*, 2 vols. Fungiflora, Oslo, Norway.

Harley, J. L., and S. E. Smith. (1983). Mycorrhizal Symbiosis. Academic Press, London.

Hudler, G. (1998). *Magical Mushrooms, Mischievous Molds*. Princeton University Press, Princeton, New Jersey.

Lincoff, G. (1981). *The Audubon Society Field Guide to North American Mushrooms*. Alfred Knopf, New York.

Raper, K. B. (1984). *The Dictyostelids*. Princeton University Press, Princeton, New Jersey.

Rippon, J. W. (1988). *Medical Mycology. The Pathogenic Fungi and the Pathogenic Actinomycetes*, 3rd ed. W. B. Saunders Company, Philadelphia.

St-Germain, G., and R. Summerbell. (1996). *Identifying Filamentous Fungi: A Clinical Laboratory Handbook*. Star Publishing Company, Belmont, California.

Stamets, P. (1993). *Growing Gourmet and Medicinal Mushrooms*. Ten Speed Press, Berkeley, California.

Volk, T. J. *Tom Volk's Fungi Webpage:* www.wisc.edu/botany/fungi/volkmyco.html

Volk, T. J. Mycological Society of America web page. www.erin.utoronto.ca/~w3msa

Volk, T. J., M. E. Kozak, and J. Krawczyk. (1997). Ecological guides to the cultivation of edible mushrooms. *Mushroom News* 45(5), 26–36.

Weber, N. S. (1988). *A Morel Hunter's Companion: A Guide to the True and False Morels of Michigan*. Two Peninsula Press, Lansing, Michigan.

GENE BANKS

Simon H. Linington and Hugh W. Pritchard
Royal Botanic Gardens

GLOSSARY

accession Usually a sample (e.g., seed lot) but may be a set of genetically related samples.

collection One or more accessions.

gene pool The genetic diversity contained within a population, species, or crop. The primary gene pool of a crop represents the biological species, the secondary gene pool includes species that can be crossed with it, allowing at least some transfer of genes, and the tertiary gene pool includes related wild species where such gene transfer involves specialized techniques.

genotype The genetic makeup of an individual.

seed lot A sample of seeds with a common harvest and post-harvest history.

wild or nondomesticated species Those species that have not been brought into regular cultivation. Many may, however, have well-known uses.

A GENE BANK is an *ex situ* (or off site) collection of genetic material held for long-term security or for ease of access. The genetic material is usually in the form of live propagules or gametes though, increasingly, pure DNA is held. A bank can also be a collection of full-grown plants representing the diversity of certain species (a field gene bank) or a set of cultures (cell, tissue, embryo, or microorganism). The widest definition might also embrace collections of plants grown primarily for purposes of education and horticulture in botanic gardens and animals represented in zoological gardens or parks.

I. ROLE WITHIN CONSERVATION STRATEGIES

Ex situ (or off site) conservation underwrites conservation measures *in situ* (on site or in the natural state) and provides centers from which genetic material (germplasm) can be accessed for research, breeding, or (re-) introduction. Nearly any *ex situ* collection might be included under the loose definition of a "gene bank." As with a financial bank, deposits are made with the intention of later withdrawal. Furthermore, while in the gene bank, safety and integrity will be paramount and value will be added through study. However, the emphasis placed on withdrawal varies considerably between gene banks. For some collections, short-term utilization is the sole motivation, though inevitably conservation will be served to some extent. In others, long-term insurance against loss *in situ* is the stronger driving force. Given the increasing risks of genetic erosion (i.e., a narrowing of the gene pool leading to species being

Encyclopedia of Biodiversity, Volume 3

condemned to extinction) as a result of human development pressures and the not inconsiderable difficulties of managing *in situ* populations, the likelihood is that many such collections will be called on in the future.

While *in situ* conservation should be seen as the primary goal of all conservationists, having a second line of defense is a prudent measure particularly if it assists in other ways. For example, by acting as a supply source, *ex situ* collections reduce the pressure of repeated sampling of wild populations. Importantly, through the use of such collections, access to national genetic resources can also be controlled. Furthermore, because research is facilitated by such collections, knowledge about both the conservation biology and socioeconomic value of wild populations is improved, thereby increasing the likelihood of survival *in situ*. It is sometimes asserted that *ex situ* conservation of certain species reduces political will to conserve the habitats in which they occur. Should this be true, it would seem essential to ensure against the loss of, at the very least, flagship species within the habitats so that they can be reintroduced. Additionally, it should be apparent that habitat complexity may limit reconstruction opportunities from *ex situ* collections. In summary and as often stated, *ex situ* conservation is complementary to that carried out *in situ* (see also the entry for "*Ex situ, in situ* conservation").

The main types of animal *ex situ* collections are zoological gardens and parks, sperm and ova banks, and DNA banks. Due to controlled breeding programs, genetic resource databanks also play a more prominent role in the management of such collections than is perhaps the case with those for plants and particularly those in botanical garden collections. Microbe collections are mainly cultures though the possibility exists for fungal spore banks.

When considering plant species, the main types of *ex situ* collections are botanic gardens specimen collections (each normally comprising many species and few individuals per species); field gene banks (usually few species, many individuals); seed, pollen, and spore gene banks (potentially many species, many individuals); *in vitro* cultures (relatively few species, many individuals); and DNA banks (potentially many species, many individuals—though the latter are rarely held separately). The collections vary in the degree of technical input required, their ability to effectively conserve inter- and intraspecific variation, their longevity, and the ease with which gene products can be obtained. Seed gene banks (in this entry abbreviated to "seed banks") nicely balance these factors and offer a very effective form of *ex situ* conservation for many plant species. For this reason and the fact that most references to gene banks relate to seed banks, this chapter will firstly consider this type of gene banking and in particular the technology.

II. SEED BANKS

A. Seed Banks Are an Effective Form of *ex Situ* Conservation

What makes seed banks such an effective *ex situ* conservation technique is that the methodology can be applied to a wide range of species in a universal and straightforward way and that large amounts of intraspecific diversity can be conserved and for long periods of time without intervention.

The fecundity of most plants means that a small seed sample can be removed from the annual seed rain with limited effect either on that population's survival or on the seed yield. One seed lot, if carefully sampled, can hold the majority of alleles found within that population. Moreover, if carefully germinated and then grown up under the same conditions, a seed lot has the potential to recreate the original population. In the meantime it can be held within a relatively small volume. For instance, a million tobacco seeds, each one with the potential to grow into a genetically distinct plant, occupy about a fifth of a liter.

Additionally, germinating seeds to obtain fully grown plants is relatively simple compared with obtaining plants from *in vitro* stored material. Plants recovered from banked seeds can also be compared with natural populations from which the material was harvested years before and which may have been subjected subsequently to environmental change (e.g., as a result of global warming).

B. The Historical Context

Seed banks might be seen as an invention of the 20th century, though the exploitation of crop seeds' ability to store in a dry state has almost certainly played a significant role in early human civilization. Early forms of seed storage probably included burying the seeds in ash or sealing them in adobe huts for the following season. By comparison, modern seed banks use controlled drying facilities and deep freeze stores and aim for storage lives of decades, centuries, or more. Modern seed banks have a lineage built on plant exploration and the development of botanic gardens that dates back nearly 5 millennia.

A collection of medicinal plants was established in

China in 2800 B.C. Three hundred years later, the Sumarians were collecting vine germplasm from Asia Minor. During the Middle Ages, the Islamic world made great strides forward in the study of plants and their properties. There followed a proliferation of gardens of medicinal plants during the late Middle Ages and Renaissance of Europe. Acquisition and exchange of crop plants gathered pace with the colonial activities of the European powers. For instance, Columbus took wheat and other seeds to the New World (Plucknett *et al.*, 1987). Following such relocation of crops and their subsequent adaptation and diversification, the current patterns of food species use were established. Current collections of crop genetic resources reflect a strong degree of interdependence across the world on genetic diversity born out of this era of colonization. For example, one study suggests that North America now relies nearly totally on, and two-thirds of developing countries now obtain more than 50% of their crop production from, crops domesticated external to their region (FAO, 1996).

European colonization spurred on the development of many of the world's botanic gardens. For instance, the French established the first tropical botanic garden at Pamplemousses in 1735 to act as a staging post for germplasm movement and to screen for use. The Royal Botanic Gardens, Kew, was established in 1759 and in turn helped develop botanic gardens elsewhere in the world (e.g., Singapore). These subsequently acted as conduits for distribution of germplasm around the world. Many famous plant collectors date from this period of European colonial expansion and include Sir Joseph Banks, Richard Spruce, and George Forest (see Plucknett *et al.*, 1987). Not surprisingly, many botanic gardens now hold important collections of both endemic and exotic nondomesticated species including those known to be used locally for medicines and timber plus many ornamentals.

Changed agricultural practice and increased pressure on the natural world during the 20th century has led to increased erosion of genetic diversity both in crops and wild species. Currently, the most important causes of genetic erosion are considered to be replacement of local varieties, land clearance (e.g., fuel-wood collection), and overexploitation; some of the greatest concern relates to diversity found in the arid and semiarid regions (FAO, 1996). In 1949 there were about 10 thousand wheat varieties in China; in the 1970s the number had reduced to a thousand. Erosion of the genetic base of crops is not new. For instance cultivated carrot color was infinitely more diverse in Medieval Europe than it is now. However, the rate of erosion has accelerated in the 20th century with the drive toward greater genetic uniformity of crop varieties. Some highly uniform varieties have been sown across huge areas. In 1983 a single variety was sown across two-thirds of the wheat fields in Bangladesh and the following year across nearly a third of all those in India. This uniformity has productivity advantages within agricultural systems but is not without risk. For instance, there was a severe shortfall in winter wheat production in the Soviet Union when a variety grown over 15 million ha did not have sufficient cold tolerance to survive the winter of 1972. Less dramatic was the abandonment in 1975 of all of the United Kingdom's white clover varieties due to susceptibility to the disease *Sclerotinia trifoliorum* (FAO, 1996).

The more modern uniform varieties have been selected from genetically heterogeneous old varieties and landraces (primitive varieties) that they have then displaced. Therein lies the conundrum. For these new varieties to be developed further they need to draw on the genetic diversity that they have displaced. Half of the production increases of the Green Revolution were brought about by the use of plant genetic resources for plant breeding.

C. The Current Status of Seed Banks for Food and Agriculture

Although the outstanding Russian genetic resource scientist Nikolai Vavilov had assembled a collection of 50,000 cereal and pulses during the 1920s and 1930s, which were maintained by an annual grow-out, the first impetus to develop modern seed banks occurred in the United States. Here it was noted in the late 1940s that less than 10% of 160,000 plant accessions in the national system since 1898 could still be found (Plucknett *et al.*, 1987). Following the establishment of U.S. regional genetic resource facilities to address this problem, the National Seed Storage Laboratory (NSSL) was established at Fort Collins, Colorado, in 1958. With about 360,000 accessions, it and the Institute of Crop Germplasm in Beijing with about 316,000 accessions are the world's largest seed banks.

A significant proportion of the accessions held in crop genetic resource collections (including seed banks) date from the 1960s onward when the extent of loss of diverse landraces was becoming apparent. Catalyzed by the Food and Agriculture Organization of the United Nations (FAO) and subsequently by the International Board for Plant Genetic Resources (IBPGR, now the International Plant Genetic Resources Institute, IPGRI), a great amount of crop genetic diversity was saved in the 1970s and 1980s through the organization of collecting missions. Not surprisingly, conservation efforts have concentrated on the 30 crop species that provide 95% of dietary energy or protein. Of these species, rice, wheat, and maize provide 26, 23, and 7%, respectively. Consequently, about 40% of the estimated 6 million accessions conserved in the 1308 national and regional collections are cereals (of which 14% are wheat). A further 15% are food legumes. Other foods such as roots/tubers, vegetables, and fruits each comprise less than 10% of the total. Because many of the species in the collections are harvested for their seeds, it is not surprising that most accessions (about 90%) are held within seed banks. The remainder are held *in vitro* or within field gene banks (see later). Of the accessions for which there are data, perhaps half are advanced cultivars or breeders' lines and a third are landraces and old cultivars. The tertiary (plus much of the secondary) gene pool within wild and weedy species comprise only 15% of the accessions (FAO, 1996). The narrow range of species is indicated by the fact that forages and forest species are considered to have minimal coverage. There is no good estimate of the number of wild species held under internationally acceptable gene bank conditions. There are probably more than 10,000 species, but almost certainly less than 20,000 species, held. The National Plant Germplasm System in the United States accounts for some 8720 species (see ten Kate and Laird, 1999). This should be set against a background of about a quarter of a million scientifically described higher plants. Obviously, the *ex situ* conservation of wild plant diversity remains a major gap that needs to be filled within the genetic resource system.

The international institutes within the Consultative Group on International Agricultural Research (CGIAR) system account for 10% of the 6 million accessions. These institutes include the following (with number of accessions, most of which are held in seed banks, shown in brackets; see FAO, 1996):

- The Centro International de Mejoramiento de Mais y Trigo (CIMMYT) in Mexico (136,637)
- The International Crops Research Institute for the Semi-Arid Tropics (ICRISAT) in India (110,478)
- The International Centre for Agricultural Research in the Dry Areas (ICARDA) in Syria (109,029)
- The International Rice Research Institute (IRRI) in the Philippines (80,646)
- The Centro Internacional de Agricultura Tropical (CIAT) in Colombia (70,940)

Just under half of the 6 million accessions are held in 12 national collections. To some degree this is a function of the early establishment of their genetic resource collections. The collections include those in Russia and the United States (noted earlier), Japan (established in 1966), Germany (1970), Canada (1970), and Brazil (1974). Of the 1308 national or regional collections currently noted by FAO (1996), only 397 within 75 countries are held in long- or medium-term seed banks. Medium-term storage might be assumed to be in the order of 10 or more years.

Whilst 496 collections are within Europe and 293 are within Asia, some areas of the world have very few such facilities. FAO (1996) noted that there is only one long-term store within the Caribbean and this is in Cuba. Relatively few of the collections covered by the FAO survey are held by private companies, though many breeders hold smaller working collections.

The remits for crop banks are very varied. Some cover wide ranges of the crops that are important at the national or regional level, others cover very specific material. For instance, facilities at Horticulture Research International in Wellesbourne (United Kingdom) and the Asian Vegetable Research and Development Centre (Taiwan) have remits for vegetable germplasm.

While the science and technology underpinning these seed banks is generally well founded (discussed later), problems such as unreliable electricity supply mean that only the nine CGIAR banks, those in 35 countries and four regional centers, meet international standards as established by FAO and IPGRI (1994). To some extent this problem is obviated by the high level of duplication within the world's 6 million accessions. It is estimated that perhaps 1 to 2 million of these are genetically unique.

The way in which seed bank and associated genetic resource work is organized both at a regional and national level varies considerably. In Europe, the European Cooperative Programme for Crop Genetic Resources (ECP/GR) works through crop networks to coordinate effort. In Southern Africa, work is coordinated through a network set up through the Southern Africa Development Corporation (SADC). At a national level, in India, genetic resources activities are formally centered around the National Bureau for Plant Genetic Resources. Within the United Kingdom, genetic resources activities are less centralized with responsibility residing with a number of institutes and their respective ministries. Advice to government is provided through the U.K. Plant Genetic Resources Group, which draws members from all of the relevant organizations.

D. Seed Banks for Wild Species

Increasingly, botanic gardens are adopting seed banking technology. Although FAO (1996) note that 150 botanic gardens (about 10% of the world total) have seed banks, relatively few of these have embraced international standards (discussed later). Those that have include the ones at the Berry Botanic Garden, Oregon (United States), the National Botanic Garden of Belgium, the Royal Botanic Gardens, Kew (Wakehurst Place), and the network of seed banks within the Spanish botanic gardens (notably Cordoba). Most botanic garden seed banks were born out of a need to more effectively manage the annual seed exchange. However, many now concentrate on the conservation of undomesticated species and interspecific variation. Additionally, several banks not in botanic gardens concentrate on wild species. These include the Threatened Flora Seed Centre established by the Department of Conservation and Land Management in Western Australia, the forest tree seed banks (e.g., the National Tree Seed Programme in Tanzania), and the seed bank of the Universidad Politecnica de Madrid. The latter bank was established in 1966 to conserve Crucifers and species from around the Mediterranean. Together with the Royal Botanic Gardens, Kew Seed Bank that has concentrated on wild species from the world's drylands, it has helped pioneer the application of seed bank technology for the conservation of wild species. Faced with the acceleration in wild plant genetic erosion and extinction, the Kew Seed Bank has taken on the ambitious objective of facilitating the conservation of seed from 10% of the higher plant species by the year 2010. Termed the Millennium Seed Bank Project, the work is a collaborative exercise involving botanical organizations from around the world. One of the aims of the Project is to maximize the efficiency of the banking work through an active seed research and technology transfer program. Other than the NSSL at Fort Collins, the project will have one of the largest seed research groups associated directly with a seed bank. Certain other botanic gardens, such as that in Rio de Janeiro, also carry out seed research work applicable to seed banking.

E. Community Seed Banks

A valuable role in seed banking is played by nongovernmental organizations (NGOs) at the community level. There is an increasing use of community seed banks within the developing world that involve donation, storage (usually on a short-term basis), and use of seeds at the local level. Examples include a set of community seed banks set up in North Eastern Brazil. The Plant Genetic Research Centre in Ethiopia is an example of a national program integrating with a community-based system (Heywood, 1995). Seed Savers in the United States and the Henry Doubleday Research Association in the United Kingdom are examples of NGOs active at the national level in the conservation of traditional vegetable varieties through their heritage seed libraries.

F. Collection Value

A financial value can be placed on biodiversity in one of two ways (see FAO, 1996):

- Direct value, or a willingness of the market to pay for it
- Indirect value, including its value as an insurance policy, the value to a product derived from the input of specific genetic material or the travel costs people would accept to collect it or to visit it for tourism

Obviously, there are other ways of considering the value of biodiversity such as its "service" value to biosystems and its value as a heritage to each human generation. However, debate has not surprisingly concentrated on its economic value and consequently the issue of access to the genetic wealth contained within seed banks is an increasingly controversial subject. Behind this debate is an attempt to reconcile the financial contribution of samples collected in one country to the economy of other countries. A fundamental concern that is often expressed is the degree of control over agricultural systems, particularly within the developing world, by multinational agrochemical/seed companies. Concern revolves around the use of the following:

- Uniform F1 hybrid varieties (where the seed, if harvested from the crop by the farmer, subsequently produces a heterogeneous crop of lower yield)
- Varieties that work best within systems when treated with certain agrochemicals
- The use of genetically modified organisms (GMOs)
- Plant breeders' rights and, increasingly, patents

Set against a backdrop of a $15 billion estimated annual turnover for the seed industry in Organization for Economic Cooperation and Development (OECD) countries, the debate has polarized on an "Industrial

North–Developing South" divide. This has led to a useful reassessment of the sharing of benefits, though inevitably there has been some over expectation of the value of individual germplasm collections. Undoubtedly, there are examples of large financial contributions to the commercial world from plant germplasm. The value of genes for high sugar content from *Lycopersicon chmielewskii* have been estimated to contribute $5 to 8 million annually to the tomato industry (see FAO, 1996). However, it is worth bearing in mind that these are the exceptions to the rule. Relatively few advances in plant breeding have been spectacular and few have resulted from the incorporation of single genes. Most have occurred from the gradual improvement of characteristics governed by many genes (polygenic characters). Of course, advanced genetic manipulation techniques may change this picture and, indeed, may have an impact on the perceived value of traditional crop genetic resources.

In contrast to the seed industry, the pharmaceutical sector is estimated to be valued at $235 billion (FAO, 1996). Included within this turnover are plant-derived drugs, many of which relate back to plants collected from developing countries. However, large financial returns from plant-derived drugs are unusual because major breakthroughs in the identification of new bioactive compounds from plants are infrequent. It is true that 34 out of the top 150 prescription drugs in the United States have a plant-based origin. This aside, despite the estimated tens of thousands of species investigated in bioprospecting programs, only 90 plant species have yielded 119 compounds considered to be important drugs (see ten Kate and Laird, 1999).

The debate has driven the establishment of agreements covering the access to and the sharing of benefits derived from plant genetic resources. Consequently, seed bank curation needs to accommodate not only national and international legislation but compliance with increasingly strict conditions attached to the material by the country or organization originally providing the material to the seed bank.

The Convention on Biological Diversity (CBD) came into force on December 29, 1993, following its opening for signature at the Rio Earth Summit in 1992. It has provided party governments with a framework within which access to genetic resources and the sharing of benefits arising from their utilization can be negotiated. Many countries have either developed, or are preparing national legislation to implement, access and benefit-sharing provisions of the CBD. Fundamental concepts that are often incorporated into national access legislation are as follows:

- The need to obtain *prior informed consent* for acquisition of genetic resources from the appropriate national authority
- The need to reach *mutually agreed terms*, which implies a negotiation leading to a form of bilateralmaterial transfer agreement specifying the terms and conditions under which material is transferred

In addition to the provisions of the CBD and implementing legislation at the national or regional level, certain seed bank collections may fall within the ambit of the International Undertaking (IU) on Plant Genetic Resources for Food and Agriculture (PGRFA). This is currently being revised in order to establish a *multilateral* access and benefit-sharing system for PGRFA that is compatible with regimes established under the CBD. It also recognizes the contribution made by indigenous farmers, in particular, to the development of plant varieties.

Seed banks, particularly the larger ones such as those within the CGIAR system (discussed previously), have responded to the emerging legal framework on access to genetic resources and benefit-sharing by reviewing the terms under which they acquire genetic resources and supply them to third parties. In many cases, seed banks have adopted the use of standard material transfer agreements that entitle the providing organization (or the country of origin of the material) to a share of any benefits arising from the utilization of the material. They may also restrict the subsequent transfer and use of the material. A pioneering worldwide project by 16 botanic gardens, which like the agricultural seed banks have traditionally supplied genetic resources to a wide range of users, has also attempted to find broadly standard terms under which material may be transferred to third parties.

Particular challenges faced by seed banks include the following:

- The acquisition of new material on terms that enable its subsequent distribution and use, so far as possible, on a common basis
- The development of a supply policy flexible enough to ensure that terms and conditions of any bilateral agreement (e.g., with the country of origin) are met, while enabling plant genetic resources to be also made available for food and agriculture under any multilateral system to be adopted in the revised IU
- The development of procedures that enable curators to clearly identify whether a proposal to acquire or supply material should be dealt with on a

multilateral or bilateral basis. For instance, due to the single source nature of many successful pharmaceutical "hits," such work may fall into the latter category

Finally, it is important to note that the benefit of the annual investment in PGFRA conservation and utilization compared with its estimated total annual cost of $1 billion (FAO, 1996) is considered to be very high. For this reason it is essential that the future operations of seed banks are not threatened by financial insecurity.

G. The Scientific Principles Underlying Seed Banking

Viable seeds of many species when maintained in a dry and cold state are capable of being germinated many years later. This capability means that the long-term *ex situ* conservation of many higher plants is a realistic possibility.

1. Seed Storage Conditions

The science of seed storage is not a new one and dates back, at least, to China in the 6th century. Advances in the quantification of seed longevity under different storage conditions were made in the second half of the 20th century through the work of Harrington in the United States and Ellis and Roberts in the United Kingdom (see review in Hong, Jenkins, *et al.*, 1998). Critical factors that determine seed longevity are the seed's moisture content, temperature, and gaseous environment; its initial viability; and its genetic background. With respect to the latter, differences between species would appear to be much greater than those within. Genetics particularly influences the relationship between seed longevity and seed moisture content. Most species produce seeds that can be dried to low moisture contents (e.g., where less than 5% of the seeds' fresh weight is water) without loss of viability. The seeds of such species are termed "orthodox." Furthermore, orthodox seed longevity increases in a predictable fashion as the moisture content is reduced. Within limits, there is a straight-line relationship of negative slope between longevity and seed moisture content (up to about 15 to 25%) when both are expressed on a logarithmic scale. More simply, Harrington's rule of thumb states that a 1% reduction in moisture content roughly doubles seed longevity. Ellis *et al.* in 1989 found that drying seeds below a certain moisture content (equating to that in equilibrium with air at about 11% relative humidity at room temperature) did not increase longevity in most species studied. However, in a few species, a decrease in longevity was noted.

Temperature also has a predictable effect on seed longevity over the range −13°C to 90°C (Dickie *et al.*, 1990) such that there is a quadratic relationship between longevity on a logarithmic scale and temperature—although an Arrhenius relationship also fitted the data well. Importantly, the relative effect of temperature on the seed longevity of eight species was identical. In its simplest form, Harrington's rule of thumb states that a 5°C reduction in temperature doubles seed longevity.

In 1980, Ellis and Roberts developed a predictive model for orthodox seed storage that incorporated the determinants of longevity (except gaseous environment that has minimal effect at the low moisture contents over which the model is applicable). Based on acceleration of the aging process in seed lots by increasing their moisture content and temperature, the model allows extrapolation of longevity at cooler and drier conditions. For instance, rice seeds at 5% moisture content and −20°C (fairly typical seed bank conditions) have a theoretical potential to survive at least 1900 years (see Hong *et al.*, 1998). Such data lend weight to occasional press reports of the germination of seeds of great antiquity, many of which cannot be confirmed due to the lack of supporting archaeological evidence. Nonetheless, there is, for example, evidence of longevity greater than 100 years at room temperature for dry wheat seeds (Steiner and Ruckenbauer, 1995). The seed storage model indicates the independence of temperature and moisture content on longevity over a wide range of conditions (Ellis *et al.*, 1995). It has been suggested that this may not necessarily be the case, particularly at very low moisture contents (Vertucci and Roos, 1990). The model also appears to be influenced by developmental stage (Hay *et al.*, 1997). These deliberations do not, however, alter the current seed bank storage recommendations (details of which are given later).

Even if these longevity predictions prove to be unattainable for all species, they do imply that, for some at least, storage in seed banks may be sufficiently long to carry the current heritage of genetic material to new eras when other technologies will be available. Predicted longevities of this magnitude also imply that the amount of regeneration of seed stocks inherent in many gene bank programs currently may be more a function of preferred working practice (e.g., regular study of collections), or poor application of the technology, rather than a failure of the technology itself.

In a study of the seed storage behavior of 6700 species, Hong *et al.* (1998), found that 91% had orthodox (desiccation tolerant) seed. The remainder had either desiccation susceptible or "recalcitrant" seed or had characteristics intermediate between the recalcitrant and orthodox types. Also certain genera (e.g., *Acer*) contain both orthodox and recalcitrant species. There is little evidence, however, that individual species produce both orthodox and recalcitrant mature seeds. It should, though, be noted that orthodox seed passes through a desiccation susceptible phase during development. The true percentage of species with orthodox seed within the plant kingdom is probably significantly less than 91% as the sample was biased toward species in trade. Although recalcitrant seed behavior is loosely associated with long-lived perennial species producing large fleshy fruits, there is great uncertainty of the number of wet tropical forest species producing such seeds. Seed storage behavior is usually assessed by quantifying the survival of seed lots of a species after drying to different moisture contents. A more rapid biochemical diagnosis of potential survival of drying is clearly of interest. Most recalcitrant seeds, including oaks and rubber, die below a value of about 40% moisture content. Because they need to be kept moist (and aerated) they have short life spans, generally of a few months, and are not suitable candidates for seed banking. Other forms of genetic conservation must be employed. The same is true for species, such as coffee, which produce "intermediate" seeds that cannot be dried much below about 10% moisture content. They also tend to be susceptible to storage at cold temperatures (−20 and 0°C). Work on other species, such as orchids, suggests the use of other subzero temperatures may be appropriate for the conservation of species sometimes included in this storage category (Pritchard *et al.*, 1999).

2. Genetic Considerations

The genetics of seed bank storage is an important issue. A criticism occasionally leveled at seed banks is that there is selection in storage. Selection can occur during collection through a biased sample (e.g., for early or late flowering genotypes). It can also occur when samples are grown out under conditions that differ dramatically from those where the seeds were harvested. There is, however, little evidence to show that, compared with, say, room conditions, the seed bank environment does other than slow down the (normally distributed) times for individuals to die. In other words there is no greater risk of selection out of any individual by the conditions applied than under natural conditions.

Another criticism is that seed lots in a seed bank are protected from evolutionary pressures that maintain fitness. The extent of this problem will depend on the generation time of the species involved and the severity of the selection pressures (at the site from which it was collected) during the time the seed lot was in storage. It might be argued that those species with short generation times will quickly adapt to all but the most extremely changed environments (i.e., the gene frequencies in the seed lot will quickly be selected to meet the new needs). Of course, maintaining the original genes sampled from within the population may have longer-term benefits of fitness for the species if pathogen attack has altered the natural population in the intervening period. An additional consideration is the prospect of genetic contamination of natural populations by chance hydridization with exotic material including genetically modified organisms. This concern is also partially alleviated by the existence of long-term stored germplasm in its original state.

H. Seed Bank Management

1. Procedure

The basic elements of the seed banking procedure (more or less in order) are as follows:

- Collection planning and permission seeking
- Seed (and pressed specimen) collecting and field data recording
- Shipment of the seeds
- Creation of a data record about the accession
- Seed cleaning (sometimes preceded by initial drying and sometimes accompanied by X-ray analysis and quantity determination)
- Main drying
- Seed moisture determination
- Initial germination test (sometimes left until after banking)
- Packaging and banking and security duplication
- Characterization (including verification of identity in the case of wild species) and evaluation (where appropriate)
- Distribution of stocks to users (through time)
- Germination retests (through time)
- Regeneration/multiplication (as required)

Although many of the well-established seed banks broadly follow these procedures, most modify them to meet their own specific needs. Some banks carry out very little field collecting and some multiply the seed sample on arrival. Costs for each stage vary considerably between seed banks. The RBG Kew Seed Bank, which has an international collecting program, estimated, in 1997, that the ratio of costs between collecting, processing, and annual maintenance for a collection was in the order of 100 : 50 : 1.

2. Seed Collecting

Set against the background of the CBD and the IU, collecting should only be carried out with the permission of the national and local authorities. Foreign collectors should work collaboratively with local scientists and clear agreements on benefit sharing should be in place. One immediately tangible benefit is the sharing of collections and the information relating to them. Other international regulations need to be adhered to. These include the Convention on International Trade in Endangered Species (CITES) and national quarantine laws.

Seed collecting methodology and genetic resource exploration have been thoroughly covered by Guarino *et al.* (1995). In most instances, random and even sampling of wild plant or crop populations is recommended, including careful note taking of the sample method (in often less than perfect conditions). Objective data recording is essential as is accurate recording of location. This latter aspect is now facilitated by the use of Global Positioning System receivers that help fix latitude, longitude, and even altitude using satellites.

Harvesting seed that is close to maturity and keeping the seed alive in the field should be paramount. Also it is essential that harvested seed is returned to the seed bank facility as securely and rapidly as possible. Prestorage conditions influence seed viability at the start of storage. Because the longevity of a species stored under specific seed bank conditions is substantially fixed, the only other key longevity factor that can be varied is initial seed storage viability. Delay in drying the seed properly or unduly harsh cleaning methods can severely reduce initial viability; in turn, this can dramatically reduce both seed longevity and seed bank efficiency. Where space is not limiting within a cold store and the longer seed lots can be kept alive, the less the unit fuel cost per species.

It is worth noting that seed lots of the same species should not be mixed either within or between years. Such action could diminish both the unique genetic makeup and reduce the longevity (through lower initial seed quality) of a seed lot.

3. Seed Cleaning

Seed lots harvested from some populations of wild species in particular contain large numbers of "empty" (or aborted) and insect-infested seeds. They may be outwardly similar to the "competent" seed and often are thus not removed during the cleaning process. Their presence can be detected by X-ray analysis or a simple cut-test applied to a subsample. Once detected, such incompetent seeds need to be allowed for when distributing seed to users and when carrying out germination tests. Although the insects may not be removed at this stage, nearly all adult and most larval forms of insects are killed by standard seed bank practice. However, insect eggs along with fungal and viral contaminants can survive seed banking. Thus, appropriate quarantine procedures need to be in place when the seed is removed from store.

4. Seed Bank Storage Standards

No two banks are the same. Traditionally, seed banks have been classified into the following categories:

- *Base* collections that are for the long-term storage of seed lots and from which seeds are not normally sent to users (though this is not always the case).
- *Active* collections from which seeds are made available to users. Often such stores are maintained under less optimal storage conditions compared to those holding base collections though this need not be the case. FAO/IPGRI (1994) recommends that storage lives of 10 to 20 years might be appropriate.

Base and active collections can occur at the same or different genetic resources centers or, as is the case in the SADC genetic resources network, the active facilities in different countries are linked to a regional base facility. In theory, the collections held within the active store need occasionally to be refreshed by the same samples held in the base store. Because active stores are likely to more frequently regenerate (discussed later) their samples, selection and genetic drift could cause a genetic divergence from the same samples held in the base store.

The FAO/IPGRI standards (1994) for base collection storage requires fresh seed to be dried within the range of 3 to 7% moisture content, packaged in a moisture-proof way, and placed at subzero (and preferably −18°C) temperatures. Although many of the world's seed banks aspire to such conditions, as has been noted, relatively few are able to emulate them. Nonetheless, the robust nature of seed bank technology means that substantial longevities are achievable even with partial fulfillment of the standards that have been discussed and using a low technology approach.

The standards set out in 1994 are important as they set the benchmark for standard operating procedures by seed banks. Consolidation of these standards and increased use of a common format for data would facilitate exchange of information between facilities. Such unification of data standards could ultimately embrace all specimen banks including those specializing in the storage of samples for environmental monitoring.

5. Seed Drying

Drying is perhaps the key to success in seed banking of orthodox-seeded species. Not only is storage prolonged in orthodox-seeded species but germination is prevented, the risks of insect and mite attack are reduced, and the seeds are protected from freezing damage. FAO (1996) found good drying facilities to be a limiting factor in a number of seed banks. Drying involves the manipulation of the water potential gradient between the inside and the outside of the seed. To all intents, the water potential is the difference between the activity of water molecules in any system compared with pure water. When the water potential of the air is lower (more negative) than that of the seed, there is a net movement of water out of the seed and drying occurs. As the water potential gradient reduces, drying slows down. When no gradient exists, the seed moisture content is said to be in equilibrium with the surrounding air conditions. Consequently, factors that influence the water potential of the air surrounding the seed are important to drying. Lowering the relative humidity or increasing the temperature of the air lowers its water potential. By keeping air moving over the seed, low humidity can be maintained by preventing moist air from accumulating around the surface of the seed. Seed size and seed depth affect rate of drying. Migration of moisture from the center of a large seed to the outside will take longer than from the center of a small seed. The same principles apply to large and small seed sacks. Shape, affecting surface to volume ratios, and seed structure will also have an effect on drying rate as can seed maturity.

It is worth noting that at any particular water potential, the seed moisture content will depend on its chemical composition. Seeds with a higher oil content will have a lower moisture content.

Most seed banks are located where the ambient relative humidity is not sufficiently low to allow the seeds to dry to the moisture content levels set out in the previous standards. High temperature gives rapid drying. However, although shade and oven drying can be used for drying, there are seed aging dangers of placing wet seeds at high temperatures and of leaving dried seeds too long under such regimes. Similarly, sun drying has risks associated with radiant heat gain. A degree of caution needs to be exercised if these methods are used. Consequently, seeds are usually placed in a drying environment where the relative humidity has been artificially lowered. Most often, this is achieved by sorption or occasionally by refrigeration drying systems. At their simplest, sorption systems can consist of a closed container into which the wet seeds are placed and the air dried by a quantity of silica gel or dried rice. Such methods require a degree of experimentation to achieve the desired results. Greater control over the extent of drying can be achieved by use of a suitable saturated salt solution (e.g., that of lithium chloride) that will maintain a set relative humidity (about 11% at room temperature) within a closed container. However, many of the larger seed banks have now adopted controlled drying rooms that allow for large samples to be dried in thin layers. Air from the chamber is dried using dryers containing lithium chloride or silica gel, cooled, and then returned. In effect, this is a closed system though most facilities allow for some fresh air intake. Ducting the air to and from the dryer on opposite sides of the chamber encourages air movement. Conditions within the chamber are those necessary to achieve the desired moisture contents at equilibrium for long-term storage. Such conditions may be 10 to 15% relative humidity at 10 to 25°C, and equilibrium is usually achieved in about one month.

6. Packaging

Maximizing orthodox seed longevity depends on drying and freezing. Cooling ambient air increases its relative humidity and hence its water potential. Placing the dried seeds directly inside a refrigerator will cause the seeds to absorb moisture. Therefore seed banks need to carefully package the dried samples. In 1996, IPGRI reviewed the types of container used in conventional long-term seed stores. Types of container used by banks include laminated aluminum foil bags, sealed steel cans, sealed glass tubes, screw-top glass bottles, and lever-top fruit preserving jars. All containers have some limitations but, carefully managed, risks of moisture ingress can be minimized especially if some monitoring system is in place. The type of container chosen will partly depend on the frequency of access required. The main concern with all is that the seal though ingress of moisture through the fabric of the material over long periods is a risk where foil is punctured or cans have poor seams. Glass offers the advantage that the contents are visible but is of course at risk from breakage. A number of banks double-pack for added security and others add a desiccant. A number of facilities dry the air in the cold store and store the collections in paper or cloth bags. Consequently, these facilities must have adequate generator backup. This is less of a problem where seeds are held in moisture-proof containers and there is a loss of electrical power. In such circumstances there is little evidence to suggest that more than a few days storage life is lost per disruption.

7. Storage Temperature

Many long-term seed banks store seed under deep freeze conditions using either purpose-built prefabricated cold rooms or domestic deep freezers. To reduce staff time at subzero temperatures, a few seed bank cold rooms, such as one at the National Institute of Agrobiological Resources (NIAR), Japan, have mechanized banking/retrieval systems. The use of such systems have implications to energy consumption by the bank.

Use of permafrost has been considered for long-term duplicate storage of seed in places such as Svalbard. Although the dependence on electricity is cut, such stores are usually unable to match the lowering of temperatures possible in conventional base storage conditions. In 1997 the Japanese-based Biological and Environmental Specimen Time (BEST) Capsule 2001 Project discussed the possibilities of long-term storage of flagship samples under Antarctic ice at −58°C (which incidentally is not sufficiently low for animal tissue preservation) or even on the dark side of the moon at −230°C.

More usually, seed storage at ultra-low temperatures (cryopreservation) is achieved using liquid nitrogen. Seed samples are normally held in polypropylene (or similar) screw-cap containers placed in the vapor phase above liquid nitrogen (about −160°C). Preferably samples of dry seed should be cooled and rewarmed at a relatively slow rate (about 10°C min^{-1}) to reduce problems of rapid thermal contraction and expansion that can cause physical injury to the seeds, such as cracking of the embryo tissue. The largest cryogenic seed bank is operated by the National Seed Storage Laboratory (United States) where there are more than 37,000 accessions stored over liquid nitrogen. More than 11,000 of these are also stored under conventional seed bank conditions at −18°C. The National Bureau for Plant Genetic Resources in New Delhi also conserves seeds under conventional and cryogenic storage conditions. Moreover, seeds of a significant number of North American and Australian wild species have been tested successfully for their tolerance of cryopreservation (Pence, 1991; Touchell and Dixon, 1993). It is worth bearing in mind when considering setting up a cryogenic facility for seed material that such low temperatures are not necessary to achieve practical periods of long-term storage (Pritchard, 1995), that the setup costs are relatively expensive, and the storage volume is less efficient than a conventional seed bank. However, these additional costs may be acceptable when creating the ultimate "base collection" for material that is in short supply, is inherently short-lived, or is of particular commercial value (unique genotypes).

8. Monitoring Seed Lot Viability

Perhaps one of the most important parameters of seed bank effectiveness is the result of germination monitoring. Germination is the preferred test for seed lot viability. Providing such information to those using the seed is helpful. Additionally, other viability tests such as vital staining using tetrazolium solution have a greater element of subjectivity about them. This staining test is, however, sometimes used to help distinguish between dead and dormant seeds among those that did not germinate under a given test regime. Two problems relate to the germination monitoring of seed bank accessions. First, because seeds are tested soon after arrival at the bank and then at regular intervals (often every 5 to 10 years) during their storage life, the tests need to be repeatable and operator independent. Second, in order to recover as many genotypes as possible represented within a seed lot, it is necessary to break seed dormancy. This can be a particular problem in seed of wild species and where the seed is freshly harvested. Key techniques include scarification of hard seed coats to facilitate water or oxygen permeation, imbibed chilling at 5 to 10°C, and later incubation at diurnal alternating temperatures with fluorescent light (i.e., rich in red light) provided only in the higher temperature phase. Tests have to be seed lot specific as most seed dormancy is not strongly genetically inherited and the form it takes depends on the conditions under which the seed matured. Once determined, the same treatments can be used during the monitoring of that seed lot through time.

Seed banks use a variety of media for germinating the seeds, such as filter paper and sterilized sand. However, 1% (w/v) plain water agar is increasingly popular especially as it reduces the risk of imbibition damage to dry seeds.

Obviously through time these monitoring tests can consume a significant proportion of each seed lot. Ideally, when the seed viability has fallen to a level where regeneration needs to take place, sufficient seeds should remain in the collection to make several attempts at growing out the collection. Because base collections may have a projected storage life of as much as 200 years, collection size needs to be large. The international standards (discussed earlier) recommend at least 1000 seeds per seed lot.

Where seed banks handle seed with unknown storage behavior, the initial germination could be delayed until after drying and banking. Survival will indicate orthodox characteristics. Few banks will have sufficient staff time to do more than brief tests prior to this on suspected recalcitrant or intermediate seeded species.

9. Duplication

One of the main advantages of seed banks is that they centralize collections of genetic material making them more easily accessed and studied. Indeed, some seed banks might be seen as some of the world's greatest plant diversity hot spots with more individuals and, in some banks, more species per square meter than anywhere else on the planet. This centralization poses a risk to all but the most carefully located and constructed facilities. Potential catastrophic loss, which of course threatens plants conserved both *ex situ* and *in situ*, mean that duplication of collections and their associated data is an important element of seed bank safety. The FAO report (1996) indicates that the level of security duplication of plant genetic resources for food and agriculture still needs to be improved and is at best uncertain.

10. Characterization and Evaluation

Characterization can vary from accurate naming of the species or subspecies represented by the collection through to more detailed recording of characters governed by genes that are little modified by environmental factors (major genes). Such information, published in the form of descriptor lists, is of great value to plant breeders wishing to narrow their choice of material from, often, vast collections. Similarly, the concept of core collections has been established to facilitate use by breeders. A core collection genetically represents a limited set of accessions of a crop gene pool with the minimum of repetition.

Increasingly, characterization is taking the form of more detailed molecular techniques such as screening by Amplified Fragment Length Polymorphism. By contrast to characterization, evaluation records data on traits such as yield that are strongly influenced by the environment in which the plants are grown. Such data are thus site and year specific and are perhaps of less use to breeders.

11. Distribution to Users

A very important element of seed bank work is to make the seed available wherever possible. In evidence of the scale of such dispatch, ten Kate and Laird (1999) quote an annual distribution of nearly 120,000 samples by the U.S. National Plant Germplasm System of which 65% are sent abroad, many requested by reference to the Germplasm Resources Information Network (GRIN) available on the Internet. Furthermore, the usage rate of crop banks by plant breeders is probably less than the rate of request from banks holding broader plant diversity collections where uses include a wide array of pure and applied research in addition to field trials. During 1994–1996, there was a 50% request rate for seeds offered through an extensive list offered by the Kew Seed Bank.

Several elements need to be considered concerning the distribution of seed to users. The recipient should be provided with accurate information about the collection and how to germinate the seeds. It should also be remembered that certain species require a symbiont for effective growth (e.g., legumes and *Rhizobium*), and that the user may need to draw on germplasm for both plant and symbiont. The seed sample must be dispatched having fulfilled all necessary plant health and, where appropriate, CITES requirements. Finally, to meet the needs of the CBD and to clarify the conditions under which the material can be used, all germplasm samples are increasingly dispatched under material supply agreements.

12. Regeneration

Seed bank accessions are grown out for the purposes of regeneration of seed stock (either when seed numbers are low or when viability has reduced), for characterization and for evaluation. Many banks have a regeneration standard below which the germination of a seed lot should not fall. This is usually set at 85%. This high value limits the risk of accumulated genetic damage that is associated with seed aging. Even though falling levels of seed germination are correlated with falling levels of field establishment, many banks have adopted lower standards. This may in part be due to the backlog of regeneration work that in some national facilities highlighted by FAO (1996) is nearly 100% of the collection. By collecting high-quality seed lots in good quantity, other banks have reduced the necessity for regeneration that can be time and labor consuming and that can have adverse effects on the genetics of the collection. Samples regenerated under conditions different from where they originated can experience selection. If too small a sample is regenerated, genetic drift may occur in which rarer alleles are lost through chance. Under some circumstances, recollection, if possible, may the more desirable option.

13. Seed Bank Design

Having considered the aspects of seed bank management, a brief consideration of seed bank design is appropriate (also see Cromarty *et al.*, 1985). The location of the bank is important from political, practical, and security aspects. Potential risks have to be considered be they earthquake, flooding, or radiation fallout. Some facilities are placed underground such as the seed bank at Krasnodar in Russia and the Millennium Seed Bank in the United Kingdom. Others such as the NSSL are located on the first floor to limit possible impact from structures above resultant from seismic activity. The size of most banks should be dictated by peak annual intake (seed drying and cleaning facilities), projected capacity before a rebuild is practical (seed storage), and annual collection maintenance (germination, field, and greenhouse facilities). Cold storage facilities vary from a few domestic deep-freezers up to large rooms such as one of 140 m^2 (with capacity for 150,000 samples) at NIAR in Japan.

III. OTHER TYPES OF GENE BANK

This section provides summaries of the current status of nonseed gene banks, starting with dry propagules (pollen and spores), which can be stored under conditions similar to those used for seeds, and covering normally hydrated tissues that can also be preserved under a different set of controlled conditions. Finally, field gene banks are covered and the role of botanical and zoological gardens is briefly mentioned.

A. Pollen

There are many practical reasons for storing pollen: to support work on allergenic responses, plant hybridization, and fertility; haploid plant production; and genetic transformation systems with isolated pollen (or gametes). Optimal storage conditions for pollen are similar to those for seeds (i.e., at low moisture content and subzero temperatures). In addition, there is some evidence (e.g., in maize and *Impatiens*) that longevity in dry storage is enhanced in anoxic atmospheres. Moreover, pollen has also been stored in a vacuum-dried state. The ease of storage though relates in part to the cellular and physiological nature of the pollen. Bicellular pollen, as found in Liliaceae, Orchidaceae, Solanaceae, and Rosaceae, generally tolerates desiccation to about 10% moisture content and is relatively long-lived. By comparison, tricellular pollen, as found in Graminae and Compositae, is relatively short-lived and is much more sensitive to desiccation. Most work on the long-term storage of pollen has focused on fruit tree or forestry species, for which 10 years storage at conventional gene bank temperatures (−20°C) is easily attainable. Pollen of at least 30 species are known to survive liquid nitrogen temperatures. Although pollen banking is evidently possible for many species, there does not appear to be any large-scale gene bank operation using such material.

B. Spores

Spores of many species of both pteridophytes and bryophytes are stable for months or years when dried, and this time can be extended with storage at cold or freezing temperatures. The longevity of short-lived (chlorophyllous) spores of some species can be extended significantly by drying and freezing in liquid nitrogen (Pence, 2000). Although there are data on fungal spore storage (e.g., work by Hong *et al.*, in 1998), the majority of fungal germplasm appears to be conserved in culture or through cryopreservation of hyphae (see later).

C. Somatic and Zygotic Embryos of Plants with Nonbankable Seeds

Nonbankable seeds can nonetheless be stored using alternative approaches. Usually, these revolve around the use of rapid, partial desiccation of embryos or embryonic axes to about 20% moisture content and subsequent transferal to liquid nitrogen temperature or the use of other subzero storage temperatures. Recovery levels may be improved by pretreatment of embryos with cryoprotectants, encapsulation of the material in alginate beads, or careful manipulation of the *in vitro* recovery environment. The embryos of more than 50 species have been successfully cryopreserved. Other parts of the plant can also be used to establish *ex situ* gene banks for species with nonbankable seeds, as described in the following section.

D. Vegetative Parts of Plants

Vegetative tissues of both pteridophytes and bryophytes can be banked for germplasm preservation (Pence, 2000). Gametophytes of many bryophyte species are naturally adapted to desiccation stress and can be cryopreserved after sufficient drying. Gametophytes of pteridophytes and some desiccation-intolerant bryophytes can also be frozen when provided with some cryoprotection, such as encapsulation in alginate beads followed by dehydration or the use of abscisic acid and the amino acid, proline, as a pre-treatment. Shoot tip freezing of sporophytes of pteridophytes has also been demonstrated. It is estimated that fewer than 200 taxa of bryophytes and pteridophytes combined are currently banked worldwide using vegetative tissues, primarily at the Cincinnati Zoo and Botanical Garden and the University of Kansas in the United States, but there is significant potential for increasing this number.

The National Seed Storage Laboratory of the U.S. Department of Agriculture (USDA) also cryopreserves about 1700 apple lines using dormant scion sections, which are retrieved by grafting (i.e., no culture of meristems). The lines are mainly from *Malus* x *domestica*, plus 10 to 15 other apple species. Some pear and cherry species (Towill and Forsline, 1999) are also banked in this way.

Apical shoot tips and other meristems/buds are the most popular vegetative materials for cryopreservation. Lines from about 50 species are now routinely cryopreserved. Initial studies used shoot tips from cold hardy, temperate zone species (apple, pear), but cryopreservation methods have been extended to tropical zone species (banana, pineapple). Successful cryopreservation depends on defining the physiological adaptation of the stock plant, the explant size and type, and its water content, the steps in the preservation process (cryoprotectant concentrations and rates of addition/removal; cooling/warming rates), and the recovery system. Two-step cooling procedures have been useful for some species, but vitrification procedures (solution-based systems and encapsulation/dehydration systems) are more favored because of the technical simplicity (Sakai, 1993). All methods are designed to reduce ice crystal growth in the specimen. Other vegetative material that has been cryopreserved using similar methodological approaches include cell suspensions and callus (more than 40 species), protoplasts (more than 10 species), and root cultures (5 species).

It should be noted that cryopreservation of vegetative germplasm overcomes the problem of genetic instability during storage as all cellular divisions and metabolic processes are stopped. In contrast, two other methods of *in vitro* preservation, normal and slow growth techniques, run the risk of genetic changes (somaclonal variation) in the conserved germplasm that may result in the loss of distinct genotypes.

Species are stored under normal growth conditions (e.g., *Coffea* at 27°C) for short-term purposes only. The explant (usually meristem or nodal cutting) is frequently transferred (subcultured) to fresh nutrient medium with the risk of microbial contamination, or loss through human error. To retard growth and hence extend the subculture interval, temperature and light intensity are reduced. For example, 0 to 5°C and 1000 Lux are generally used for cold tolerant species, and 15 to 22°C and reduced light intensity for tropical species. Alternatively, growth can be slowed down by the addition to the medium of chemicals to induce mild osmotic stress (e.g., mannitol) or hormonal retardants (paclobutrazol, abscisic acid). Also, maintenance of tissue under reduced oxygen (e.g., under mineral oil or liquid medium) slows growth. Under the appropriate conditions subculture intervals can be extended to one year or longer. The slow growth technique is now routinely used for the medium-term conservation of a number of species such as banana, potato, yam, cassava, and strawberry.

Although *in vitro* culture without cryopreservation poses considerable threat of genetic drift, the propagation of plant material in an aseptic environment ensures the production of disease-free stock material, which is readily accessible internationally because it satisfies most country quarantine requirements. Undoubtedly, *in vitro* culture is a valuable complementary approach to field conservation and is particularly useful when applied to species that are predominately propagated vegetatively (banana, potato, and pear), produce non-bankable or highly heterozygous seeds, and have a particular gene combination (i.e., elite genotypes; see Ashmore, 1997). The impact of these positive features of the technique is such that FAO estimates that 37,600 accessions of plant material (vegetative and embryos) are conserved *in vitro* (including cryopreservation) worldwide.

E. Animal Germplasm Samples

A majority of *ex situ* animal germplasm is maintained in zoological gardens and institutes around the world. The NIAR in Japan holds 621 accessions of animal germplasm, including silk worms in the living state. More than 100 of these accessions, mostly sperm, are cryopreserved (for a general methodology, see the discussion presented later). Similarly, the main gene bank methodology for ova is cryopreservation. At present, however, procedures for sperm and ova preservation are not well developed for wild species, even though a number of reported successes with artificial insemination and frozen semen can be found in the literature, especially for ungulates such as deer and antelopes. The concept of gamete rescue from tissues has considerable value for spermatozoa, where epididymal spermatozoa are readily obtainable post mortem and can be frozen using glycerol as a reasonably standard cryoprotectant. Oocyte cryopreservation has only been achieved in the hamster, rat, rabbit, and cow and is therefore not a practical proposition at present. Interest in freezing ovarian tissue, and then culturing follicles and oocytes by various methods after thawing, has recently been resurrected and progress has included the birth of a lamb originating from ovarian tissue autotransplanted into the donor-recipient after freezing and thawing. In another recent study, isolated rat spematogenic cells were transferred to a mouse testis, where they displayed the ability to develop into spermatozoa. The testicular cells were frozen-thawed prior to transplantation and development, and thus there may be some merit in exploring the cryopreservation of testicular cell suspensions as an alternative or adjunct to the preservation of spermatozoa. Cell suspension from genetically important animals could be used to populate the testes of common species, thus permitting the eventual harvesting of spermatozoa. It should be stressed that this technique is still only in its infancy.

Unlike animal species in which reproductive cells and tissues are stored to conserve the gene pool, in humans they are stored for the use of the couple/woman electing to have them cryopreserved. Some embryos are subsequently donated for the treatment of others. Storage is limited by law in some countries.

The storage of embryos is common practice in *in vitro* fertilization (IVF) clinics worldwide, with about 70 offering the service in the United Kingdom alone. Oocyte cryopreservation is poorly developed and few clinics store them other than for research. An increasing number are offering tissue storage.

Despite sporadic reports of cryopreservation using nonequilibrium rate cooling, the vast majority of clinics prefer conventional slow cooling rate procedures with the samples loaded into "straws" (embryos and oocytes) or plastic vials (ovary). The first step involves the addition of the cryoprotectant [about 1.5 M; 1–2 propanediol ± sucrose (pronucleate and early cleavage stage embryos), dimethyl sulfoxide (4–8 cell embryos, and ovary) or glycerol (blastocyst stage embryos)]. Ice formation is then induced at -5 to -7°C and the samples are cooled further at rates of about 0.3 to 0.5°C per minute to various subzero temperatures before storage in liquid nitrogen. After warming at appropriate rates (of about 20 to more than 300°C per minute), the samples are returned to isotonic conditions stepwise, with or without the addition of sucrose to the diluent.

F. Microorganisms

The most widely applicable preservation method for the preservation of microorganisms that retains viability and stability is cryopreservation. However, for convenience and ease of transport freeze-drying is preferred for most bacteria, viruses, and sporulating fungi (see Hunter-Cevera and Belt, 1996).

There are currently 497 collections from 60 countries maintaining microorganisms registered with the World Data Centre for Microorganisms in Japan (see Sugawara and Miyazaki, 1999). Their on-line database lists species held and the expertise and services provided by the collections along with contact addresses and links to collection websites. There are about 11,500 species held. Over 25% of the strains are held by 5 of the 497 collections (see Table I).

G. DNA Banks

DNA banks have been established in several places worldwide (Adams, 1997), the largest (more than 140,000 clones) being for plant and animal material at the NIAR in Japan. This collection is mainly constituted of rice clones (about 36,000 comprising Random Fragment Length Polymorphisms, cDNA, YAC) and pig clones (about 106,000 comprising cosmids, BAC, cDNA). The most diverse DNA bank for plants is at the Royal Botanic Gardens (RBG) Kew, which currently holds more than 10,000 DNA samples from a wide range of species. Standards of quality of preservation differ between banks and some, such as those at the Missouri Botanical Gardens (St. Louis, Missouri, in the United States) and several zoos, store only frozen tissue. Others extract DNA and purify it to varying degrees. Purposes for these banks differ as well, with some established to hold samples of a particular country or region with the intent of using these in conservation genetic studies, whereas others focus more on taxonomic and systematic studies, such as that at RBG Kew. Most banks are prepared to consider sharing aliquots of DNA or small samples of tissue with researchers at other institutions.

H. Field Gene Banks

Field gene banks are *ex situ* collections of mainly agricultural or forest species. They should be contrasted with what might be termed "farm gene banks" where crop germplasm is, in effect, conserved *in situ* by the farmer. Field gene banks normally comprise considerably more individuals per accession than is the case in

TABLE I

Collections Maintaining Microorganisms

Collection	Location	Number of strains	Type of material
Agricultural Research Service Culture Collection, USDA	Peoria, IL (US)	78,010	Algae, bacteria, fungi, yeasts, actiomycetes
American Type Culture Collection (ATCC)	Rockville, MA (US)	53,615	Algae, bacteria, fungi, yeasts, protozoa, cell lines, hybridomas, viruses, vectors, plasmids, phages
CABI Bioscience UK Centre (formerly (IMI)	Egham, UK	21,000	Fungi, bacteria, yeasts
Centraalbureau voor Schimmelcultures (CBS)	Baarn, The Netherlands	41,300	Fungi, yeasts, lichens, plasmids
Culture Collection, University of Goteborg (CCUG)	Goteburg, Sweden	28,100	Bacteria, fungi, yeasts

botanic gardens. Their particular use is for the conservation and utilization of species with the following traits:

- Have nonbankable seeds
- Have long life cycles where growing up material for regular study from a seed collection is impractical
- Are normally vegetatively propagated

Essentially, they are not a new idea as the Kayapo people of Brazil maintain germplasm collections of tuberous plants in hillside gardens protected from flood (see Plucknett *et al.*, 1987). Most countries have at least one field gene bank and FAO (1996) estimate some 527,000 accessions are conserved in this way worldwide. Examples of field gene banks include the National Fruit Collection in the United Kindgom, one for cassava at CIAT, Colombia, one for sugarcane at the Centro Nacional de Pesquisa de Recursos Geneticos e Biotecnologia (CENARGEN), Brazil, and the potato collection at the Centro Internacional de la Papa (CIP) in Peru. Such facilities are considered by FAO to be particularly important in small island developing states. It should also be noted that the Nordic Gene Bank, which operates on a regional basis, unlike many seed banks, includes within its remit the *in situ* conservation of wild crop relatives.

While they offer the opportunity for characterization and evaluation, such collections are labor intensive and are susceptible to catastrophic events. For example, a field collection of yams was lost in St. Lucia during 1994 as the result of cyclone damage. This is one of the reasons that a number of field gene bank collections are now backed up *in vitro*.

I. Botanical and Zoological Gardens

Botanical and zoological gardens may be seen as types of gene bank with relatively few individuals per accession. While zoos have embraced careful breeding programs that help maximize the genetic value of the limited collections across the world, this is much less true of collections in botanical gardens. Here species may be represented within the "botanical gardens flora" by a single genotype such is the clonal exchange of material. See also "The Role of Botanical Gardens" and "Zoos and Zoological Parks."

Acknowledgments

The authors would particularly wish to acknowledge the following for their assistance in providing information: Mr. M. Ambrose (John Innes Centre, United Kingdom); Dr. D. Astley (HRI, United Kingdom); Dr. M. Fay, Mr. R. Smith and Mr. M. Way (RBG Kew, United Kingdom); Dr. W. Holt (Institute of Zoology, London, United Kingdom); Dr. S. Miyazaki and Mr. A. Yamamoto (NIAR, Japan); Dr. B. Panis (Leuven, Belgium); Dr. V. C. Pence (CREW, Cincinatti, United States); Dr. D. Smith (CABI Bioscience, United Kingdom); Dr. S-H Tan (CBS, Netherlands); Dr. L. Towill and Dr. C. Walters (NSSL, United States); Dr. M. Wood (London, United Kingdom); and Dr. X-Y. Yang (XTBG, Peoples Republic of China).

See Also the Following Articles

BREEDING OF ANIMALS • BREEDING OF PLANTS • CAPTIVE BREEDING AND REINTRODUCTION • CROP IMPROVEMENT AND BIODIVERSITY • EX SITU, IN SITU CONSERVATION • GENETIC DIVERSITY • ZOOS AND ZOOLOGICAL PARKS

Bibliography

Adams, R. P. (1997). Conservation of DNA: DNA banking. In *Biotechnology and Plant Genetic Resources: Conservation and Use.* (J. A. Callow, B. V. Ford-Lloyd, and H. J. Newbury, Eds.), pp 163–174. CAB International. Wallingford.

Ashmore, S. E. (1997). *Status Report on the Development and Application of in Vitro Techniques for the Conservation and Use of Plant Genetic Resources.* IPGRI. Rome.

Cromarty, A. S., Ellis, R. H., and Roberts, E. H. (1985). *The Design of Seed Storage Facilities for Genetic Conservation.* IBPGR, Rome.

Dickie, J. B., Ellis, R. H., Kraak, H. L., Ryder, K., and Tompsett, P. B. (1990). Temperature and seed storage longevity. *Annals of Botany* 65, 197–204.

Ellis, R. H., Hong, T. D., and Roberts, E. H. (1995). Survival and vigour of lettuce (*Lactuca sativa* L.) and sunflower (*Helianthus annuus* L.) seeds stored at low and very-low moisture contents. *Annals of Botany* 76, 521–534.

FAO (Food and Agriculture Organization of the United Nations). (1996). *The State of the World's Plant Genetic Resources for Food and Agriculture.* Rome.

FAO/IPGRI. (1994). *Genebank Standards.* Rome.

Guarino, L., Ramantha Rao, V., and Reid, R. (Eds.) (1995). *Collecting Plant Diversity. Technical Guidelines.* CAB International.

Hay, F. R., Probert, R. J., and Smith, R. D. (1997). The effect of maturity on the moisture relations of seed longevity in foxglove (*Digitalis purpurea* L.). *Seed Science Research* 7, 341–349.

Heywood, V. H. (Ed.) (1995). *Global Biodiversity Assessment.* Published for the United Nations Environment Programme by Cambridge University Press. Cambridge.

Hong, T. D., Jenkins, N. E., Ellis, R. H., and Moore, D. (1998). Limits to the negative logarithmic relationship between moisture content and longevity in conidia of *Metarhizium flavoviride. Annals of Botany* 81, 625–630.

Hong, T. D., Linington, S., and Ellis, R. H. (1998). *Compendium of Information on Seed Storage Behaviour.* 2 volumes. Royal Botanic Gardens, Kew, United Kingdom in collaboration with the University of Reading and IPGRI.

Hunter-Cevera, J. C., and Belt, A. (Eds.) (1996). Preservation and Maintenance of Cultures used in Biotechnology and Industry. Academic Press San Diego, CA.

Pence, V. C. (1991). Cryopreservation of seeds of Ohio native plants and related species. *Seed Science and Technology* 19, 235–251.

Pence, V. C. (2000). Ex situ conservation methods for bryophytes and pteridophytes. In *Saving the Pieces: The Value, Limits and Practice of Offsite Plant Conservation.* (K. Havens, E. Guerrant, and M. Maunder, Eds.), (in preparation).

Plucknett, D. L., Smith, N. J. H., Williams, J. T., and Anishetty, N. M. (1987). *Gene Banks and the World's Food.* Princeton University Press, Princeton, NJ.

Pritchard, H. W. (1995). Seed cryopreservation. In *Methods in Molecular Biology Volume 38: Cryopreservation and Freeze-Drying Protocols* (M. R. McLellan, and J. G. Day, Eds.), pp 133–144. Humana Press, Totowa, NJ.

Pritchard, H. W., Poynter, A. L. C., and Seaton, P. T. (1999). Interspecific variation in orchid seed longevity in relation to ultra-dry storage and cryopresevation. *Lindleyana.* 14(2), 92–101.

Sakai, A. (1993). strategies for survival of plant cultured cells and meristems cooled to −196°C. In *Cryopreservation of Plant Genetic Resources,* pp. 1–16. Japan International Cooperation Agency.

Steiner, A. M., and Ruckenbauer, P. (1995). Germination of 110-year-old cereal and weed seeds, the Vienna Sample of 1877. Verification of effective ultra-dry storage at ambient temperature. *Seed Science Research* 5, 195–199.

Sugawara, H., Ma, J., and Miyazaki, S. (Eds.) (1999). *World Directory of Collections of Cultures of Microorganisms.* 5th edition. WFCC World Data Center on Microorganisms, Japan.

ten Kate, K., and Laird, S. A. (1999). *The Commercial Use of Biodiversity: Access to Genetic Resources and Benefit-Sharing.* Earthscan Publications.

Touchell, D. H., and Dixon, K. W. (1994). Cryopreservation for seedbanking of Australian species. *Annals of Botany* 74, 541–546.

Towill, L. E., and Forsline, P. L. (1999). Cryopreservation of sour cherry (*Prunus cerasus* L.) using a dormant vegetative bud method. *CryoLetters* 20, 215–222.

Vertucci, C. W., and Roos, E. E. (1990). Theoretical basis of protocols for seed storage. *Plant Physiology* 94, 1019–1023.

GENES, DESCRIPTION OF

Michael F. Antolin and William C. Black IV
Colorado State University

GLOSSARY

allele Variant of a gene that can vary at the nucleotide level with or without affecting phenotypic expression.

chromatin Complex of DNA and associated proteins that make up chromosomes of eukaryotes.

chromosome Threadlike structure that includes DNA and proteins (containing genes arranged in a linear sequence along the thread), which can be visualized when condensed during cell division.

DNA Deoxyribonucleic acid, the molecule of inheritance that stores genetic information that is passed from one cell to another and from one generation to succeeding generations; it is composed of four nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T).

eukaryotes Organisms with genetic material organized into chromosomes that are contained within a membrane-bound nucleus in the cell; eukaryotic cells undergo mitosis and meiosis during cell division, ensuring the equal division of chromosomes among daughter cells.

exons Set of segments of interrupted genes that remain after cutting and splicing of messenger RNA, and that include the parts of the gene that are translated into proteins.

genetic code Language that specifies how DNA will be translated into protein sequences by means of three-nucleotide "words" (codons) that specify the 20 amino acids and regulators of transcription (start and stop codons).

heterochromatin Regions of chromosomes that do not include coding DNA, generally make up the structure of chromosomes, and always remain condensed during a cell's life cycle.

introns Segments of interrupted genes that are removed after transcription and before translation of messenger RNA to proteins.

locus A precise location in the genome, whether a gene is found there or not; formerly this term was used interchangeably with gene, but the definition has become more specific in the era of molecular genetics.

nucleotide Subunit of DNA and RNA composed of a ringed five-carbon sugar, a ringed nitrogen-rich base, and phosphates; nucleotides are often referred to as base pairs because individual types form complementary hydrogen bonds (G-C and A-T in DNA, G-C and A-U in RNA) to make double-stranded DNA and RNA molecules.

organelles Membrane-bound structures within eukaryotic cells that carry out specific functions and that may contain their own DNA; typical organelles are mitochondria and chloroplasts.

post-transcriptional modification Cutting and splicing of mRNA in eukaryotic cells, in some cases to produce alternative proteins with different structural or regulatory properties (e.g., sex determination in *Drosophila*).

prokaryotes Single-celled organisms defined by having their DNA arranged as a circular molecule not contained within a nucleus, and which reproduce by simple fission.

repetitive DNA Regions of DNA that include the same DNA sequence repeated up to several hundred or thousand times; regions with repeated segments that involve only 2–5 base pairs of DNA are called microsatellites.

RNA Ribonucleic acid, composed of nucleotides like DNA, but differing from DNA in that the base uracil (U) in RNA replaces thymine (T) and single-stranded RNA molecules form important structural and regulatory parts of cells.

transposable elements Fragments of DNA containing genes that provide the ability for the DNA fragment to change its location in the genome.

UNDERSTANDING THE MECHANISMS OF HEREDITY IS A KEY TO THE STUDY OF BIODIVERSITY FOR AT LEAST THREE REASONS. First, heritable genetic variation provides the material for evolution—descent with modification—including adaptation within populations, diversification of lineages, and the formation of new species. Second, standing genetic variation within populations may reflect the recent history of populations and indicate prospects for change in the future. Third, the variation in molecular genetic markers can be used to track the relationships among living groups of organisms and the taxonomic status of individual populations. Genes are defined as sequences of DNA that specify cell structure, including proteins and several types of RNA. Genes also code for enzymes and proteins that bind to DNA or RNA in order to control gene expression and other physiological processes in cells. In former times, the term "locus" was used interchangeably with "gene." Today, locus simply means a particular place in the genome, whether the DNA at that location codes for a gene product or not. Locus may still be used to describe neutral molecular markers; allelic variation at the DNA level is used to study population genetics, relatedness within populations, parentage, and genome mapping.

I. INTRODUCTION

Genes and genomes are both complex and continually modified. Organisms share a common biochemical mechanism of heredity, but at the same time display characteristic differences that allow us to make hypotheses about the history of life. Genetic variation is pervasive in Nature and continually arises by the inevitable and inexorable process of mutation, both on the level of individual nucleotides within DNA molecules and on the level of chromosomal gene rearrangements. One of the most significant (and initially unappreciated) discoveries in genetics in this century is that genomes can be altered by transposable elements. The original discovery of transposable elements was in maize by Barbara McClintock, who in the 1930s correlated regular breaks in chromosomes with the deactivation of purple pigment genes in corn seeds. Reactivation of pigment genes was also associated with the same breaks appearing at other places in the genome. We now know these as "jumping genes," which carry with them the genes that control their ability to duplicate and reinsert themselves into multiple places within genomes, sometimes carrying other genes with them, sometimes disrupting genes at their insertion sites. Dozens of transposable elements are now known, and more continue to be discovered in an ever-widening sample of organisms (Li, 1997). As a consequence, it is no longer possible to view genomes as static entities in which genes are aligned in precise locations that remain stable over long periods of time.

This article considers genes in the context of the entire genomes of diverse organisms. In prokaryotic organisms (bacteria, blue-green algae), genomes are organized as naked circular DNA molecules within the cytoplasm. In eukaryotic organisms, most genetic material is arranged into chromosomes—long, linear arrays of DNA bound with proteins—found within a membrane-enveloped nucleus. Eukaryotic organisms also have genes in separate prokaryotelike genomes inside cell organelles such as mitochondria (plants and animals) and chloroplasts (the photosynthetic organelles of plants). All genomes contain regions of noncoding single-copy DNA and repetitive DNA. Population genetic and phylogenetic studies depend on a variety of molecular markers derived from noncoding parts of genomes that are presumably neutral (i.e., not under selection), and Avise (1994) provides an in-depth overview of the use of molecular markers in the study of evolution.

II. STRUCTURE OF DNA AND RNA

What follows should serve as a review for biologists who do not regularly converse with the world of molecular genetics. A full description of genes is not possible in the space of a few pages; those seeking more detail should consult other sources (e.g., Hartl and Jones, 1998; Lewin, 1997). Molecular genetics is a mushrooming field, but the salient features of genes are universal, and genes and the genetic code can be understood on an intuitive level without extensive knowledge of biochemistry. In general, genes are carriers of information working in a nested series from individual nucleotides, to relatively large RNA molecules and enzymes composed of hundreds of nucleotides or amino acids, to enormous cell-structure proteins comprising thousands of amino acids folded into complex three-dimensional shapes.

First and foremost in this hierarchy of information are nucleic acids, DNA and RNA, which are the molecules of heredity for most life-forms (one notable exception being prions, which cause the famed "mad-cow disease" and a number of neurological disorders in humans; see the review by Prusiner and Scott, 1997). Both DNA and RNA are made up of four nucleotide subunits linked together to form long chains (Fig. 1). Nucleotides themselves are made up of three parts: ringed sugars, phosphates, and ringed nitrogen-and-carbon bases. Nucleotides are joined to each other to form chains by bonds between phosphates and sugars. Nucleotides are complementary to each other in terms of the numbers of hydrogen bonds they form between nitrogenous bases: A (adenine) and T (thymine) make two bonds, G (guanine) and C (cytosine) make three bonds. Two complementary chains intertwine to form double-stranded DNA or RNA by hydrogen bonding between bases. Double-stranded DNA molecules take on the classic "double-helix" shape, where two chains of nucleotides joined by hydrogen bonds spontaneously form twists of ~36° so that ~10 base pairs make a complete turn of 360°. DNA can take on a number of

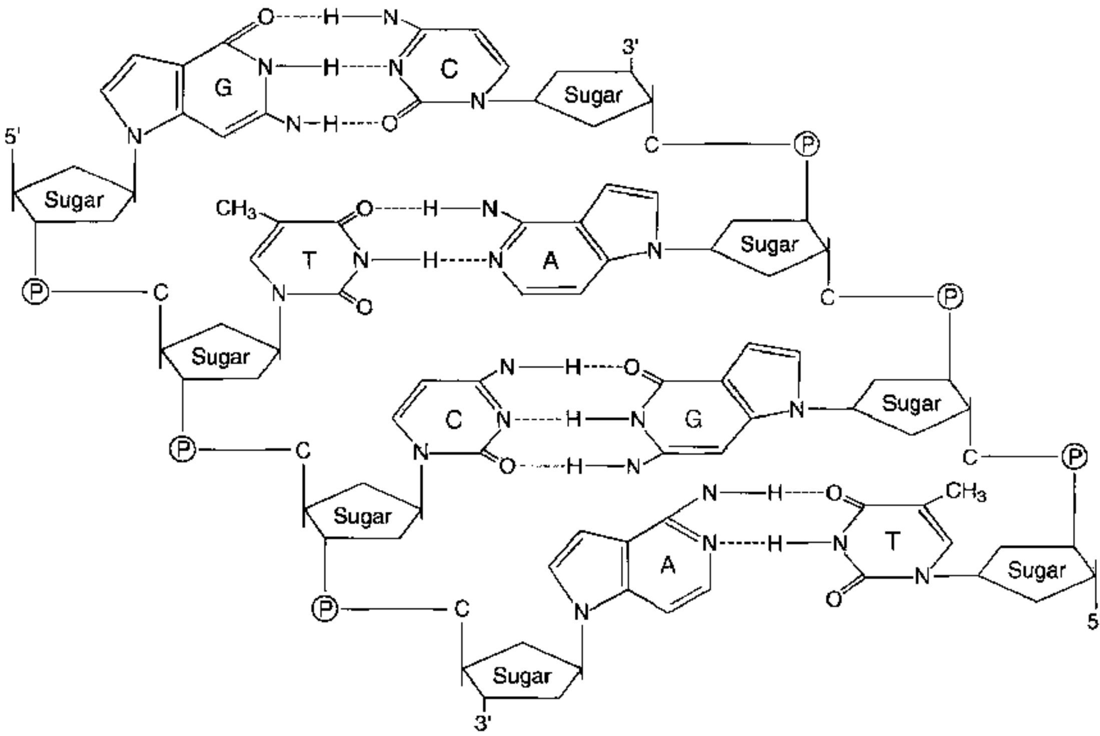


FIGURE 1 DNA (and RNA) are composed of four nucleotide subunits—A, C, G, and T (U replaces T in RNA)—each covalently bonded to a sugar, with sugars linked to each other by a phosphate bond (circled P) between the 3′ and 5′ carbon atoms of the adjacent sugars. Complementary strands of DNA are held together by hydrogen bonds (shown as broken lines). Of the two strands, the lower is in the 5′ to 3′ direction, showing how nucleic acid sequences are usually written (GTCA).

other forms, such as supercoiling structures, depending on conditions such as temperature, pH, and the exact sequence of nucleotides within the molecule.

RNA also can form a double helix, but one with a broader diameter. An additional feature of RNA is that single-stranded molecules include complementary sequences that allow the single-stranded RNAs to fold back on themselves, forming double-stranded helical stems and other three-dimensional grooves, knobs, and loops (Gutell, 1996). Three-dimensional structure is a vital characteristic of the long ribosomal RNA (rRNA) that make up subunits of ribosomes and the shorter transfer RNA (tRNA) molecules that are part of the mechanisms of translation of messenger RNA (mRNA) into protein.

III. GENES

The number of genes within organisms varies tremendously among viruses, bacteria, and multicellular eukaryotic organisms. For instance, it is estimated that humans have approximately 50,000 genes, the fruit fly *Drosophila melanogaster* has 13,600, the nematode worm *Caenorhabditis elegans* has 19,000, the mustard *Arabidopsis thaliana* has 25,000, the yeast *Saccharomyces cerevisiae* has 6000, the bacterium *Escherichia coli* has 4286, and the HIV (human immunodeficiency virus) RNA retrovirus has only 9. For the organisms whose entire genomes and all nucleotide sequences have been characterized (yeast, worm, *E. coli*, HIV), the number of genes is known with accuracy. For others, the numbers represent estimates based on incomplete mapping studies. For example, some estimates of the number of genes in humans are as high as 140,000 genes, but the upper limit is usually given as 80,000 genes.

The flow of information within a cell generally follows a pathway, with genes coded in DNA transcribed into three types of RNA, followed by translation of the nucleotide sequence in messenger RNA into chains of amino acids (proteins and enzymes). This process of information transfer, which has been called the "central dogma of molecular biology," is depicted in Fig. 2. Two kinds of genes result from the transfer of information from DNA to proteins. First, genes may encode enzymes such as DNA polymerases and protein transcription factors that control the physiological and biochemical processes within a cell. The second type of gene encodes the physical structures of the cell, including cytoskeletal proteins like tubulin and actin that affect cell shape, and subunits of ribosomal RNA that make up the machinery of protein translation. The majority of genes are expressed as proteins and enzymes, and most of these are involved in cellular regulation rather than cell structure.

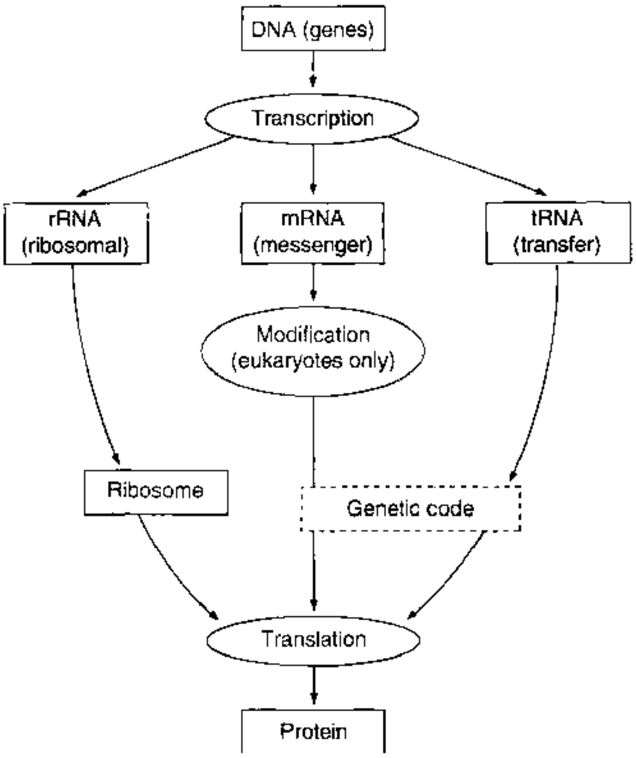


FIGURE 2 The flow of information from DNA to proteins. DNA is transcribed into three types of RNA, which then combine to read the genetic code and translate the information into proteins (or enzymes).

Protein and enzyme coding is specified by the genetic code (Table I), which contains the language of protein synthesis. The genetic code has a series of three-base codons that specify 20 amino acids as well as codons for "start" and "stop." This code is almost universal, although there are a number of differences between nuclear and mitochondrial DNA in eukaryotes and there are unique codons in a few bacteria and Protista (Li, 1997). The code functions by specifying three-base-pair codons of mRNA that match three-base-pair anticodons found on a loop of the transfer RNA. Each tRNA is specific to a particular amino acid. The complex dynamic molecular structure involved in translation of the genetic code contains mRNA, ribosomes (rRNA plus proteins), and tRNAs.

Because DNA contains specific information, transcription and translation must proceed in only one di-

TABLE I

The Genetic Code That Shows How Twenty Amino Acids Are Specified by Three-Base-Pair Codons[a]

First Position (5′ end)	Second position				Third position (3′ end)
	U	C	A	G	
U	UUU	UCU	UAU	UGU	U
	UUC PHE	UCC	UAC TYR	UGC CYS	C
	UUA	UCA SER	**UAA**	**UGA STOP**	A
	UUG LEU	UCG	**UAG STOP**	**UGG TRP**	G
C	CUU	CCU	CAU	CGU	U
	CUC	CCC	CAC HIS	CGC	C
	CUA LEU	CCA PRO	CAA	CGA ARG	A
	CUG	CCG	CAG GLN	CGG	G
A	AUU	ACU	AAU	AGU	U
	AUC ILE	ACC	AAC ASN	AGC SER	C
	AUA	ACA THR	AAA	AGA	A
	AUG MET	ACG	AAG LYS	AGG ARG	G
G	GUU	GCU	GAU	GGU	U
	GUC	GCC	GAC ASP	GGC	C
	GUA VAL	GCA ALA	GAA	GGA GLY	A
	GUG	GCG	GAG GLU	GGG	G

[a] The three stop condons (UAA, UAG, UGA) and the start condon for methionine (AUG) are shown in bold. Most amino acids are specified by more than a single codon, with several (arginine, leucine, serine) specified by six codons. There is a weakly positive general relationship between the frequency of incorporation of an amino acid into proteins and the number of codons specifying the amino acid. In most cases, a change in the third position in a codon does not result in a change in the amino acid because most tRNA anticodons will still bind to mRNA codons if the first two positions on the mRNA match the tRNA. This is the so-called wooble effect and redundancy in the genetic code.

rection for the information to be communicated accurately. This is analogous to the convention in most written Indo-European languages, in which words form sentences that are read on a page from left to right. The direction along any strand of DNA or RNA can be identified by which carbon atoms of the ringed-sugar-and-phosphate backbone are bonded together. Carbons in the ribose sugar are numbered 1 to 5, and bonds occur between a phosphate attached to the 5′ carbon of one sugar with a hydroxyl group attached to the 3′ carbon of the next. Consequently, a directionality can be specified along each nucleic acid molecule, and the genetic code must follow this direction for the language of transcription and translation to be faithfully transmitted. For instance, RNA polymerases transcribe by adding new nucleotides to the 3′ end of a growing RNA molecule, and so create the RNA in the 5′ to 3′ direction. However, the transcription is from a complementary DNA chain that is read in the 3′ to 5′ direction. In general, reading frames of DNA are written by placing the 5′ end at the left, and protein sequences are written in the same left-to-right direction specified by the nucleotide sequence (Table II).

IV. PROKARYOTIC GENOMES

The genomes of most prokaryotes (bacteria and blue-green algae) are composed of one double-stranded circular DNA molecule attached to a central core of proteins in a series of supercoiled loops emanating from the center like the cotton fibers of a dust mop. Although the nucleoid "chromosome" is not enclosed within a nuclear membrane, it tends to be found in a particular part of the cytoplasm. An additional feature of prokaryotes is their ability to acquire smaller circular DNA molecules, called plasmids, in their cytoplasm. Plasmids may contain numerous genes, including genes found in the nucleoid. Transfer of plasmids into and out of bacterial cells is the major feature that makes them vital

TABLE II

Types of Chromosomal and Gene Mutations[a]

A. Chromosomal Mutations		
	DNA sequence (of a chromosome)	Type of mutation
Original sequence	AGTTCGTAGGGTACCTGATCGACG	
Mutant sequences	AGTTCGTACCTGATCGACG	Deletion of AGGGT
	AGTTCGTAGGGTACCTGATCATGAGCGACG	Insertion of CATGAG
	AGTTCGTAGGTCCATGGATCGACG	Inversion of GTACCT
	GACGGTAGGGTACCTGATCAGTTC	Translocation of AGTTC and GACG
	AGTTCGTAGGGTACCTGACCTGATCGACG	Duplication of CCTGA
B. Gene Mutations		
	Three-base-pair codons (top line), and the proteins specified (lower line)	Type of mutation
Original gene	AUG UGG UGU GAG UAC AUU CGA GAG AAG UAG met-trip-cys-glu-tyr-ile-arg-glu-lys-stop	
	Mutations at the third codon	
Mutant genes	AUG UGG UGC GAG UAC AUU CGA GAG AAG UAG met-trp-cys-glu-tyr-ile-arg-glu-lys-stop	Synonymous (no amino acid change)
	AUG UGG UCU GAG UAC AUU CGA GAG AAG UAG met-trp-ser-glu-tyr-ile-arg-glu-lys-stop	Mis-sense (amino acid change)
	AUG UGG UGA GAG UAA AUU CGA GAG AAG UAG met-trp-stop	Non-sense (mutation to a stop codon)
	AUG UGG UUG AGA GUA AAU UCG AGA GAA GUA G met-trp-leu-arg-val-asn-ser-arg-glu-val-	Non-sense (insertion and shift of reading frame)
	AUG UGG GAG AGU AAA UUC GAG AGA AGU AG met-trp-glu-ser-lys-phe-glu-arg-ser-	Non-sense (deletion and shift of reading frame)

[a] Very small fragments are illustrated; chromosomal mutation usually affect large fragments. After Li (1997).

to DNA-based biotechnology and is responsible for the rapid transfer of genes for traits like antibiotic resistance between bacterial strains.

Genes in prokaryotes are collinear and uninterrupted, in that one to several genes are expressed in the order in which they follow a promoter. Transcription of mRNA leads directly to translation with little modification (Fig. 3). This arrangement of several genes being expressed as a group is sometimes referred to as polycistronic, and it is restricted to prokaryotic and related genomes (e.g., organelles; see Section VI). The control of gene expression in prokaryotes is through either positive or negative regulation. Under negative regulation, like that found in the tryptophan operon of *E. coli*, a repressor protein combines with a tryptophan molecule and then binds to the repressor region of the gene, turning off transcription when tryptophan becomes common within the cell. Under positive control, a transcription factor must bind with the regulatory site to initiate transcription. This is part of the *lac* operon of *E. coli*, one of the earliest metabolic pathways described at the molecular level by Jacques Monod and colleagues at the University of Paris. The *lac* operon is mainly under negative control of a repressor that is removed when lactose is present in the cell. However, transcription also depends on a second positive control on the promoter, where a second protein must bind for transcription of RNA. This double control allows the

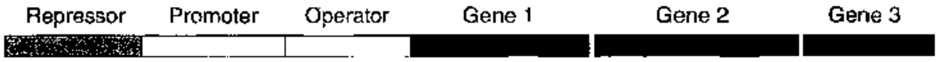


FIGURE 3 General structure of a prokaryotic gene, in which a repressor, promoter, and operator act together to control expression of three genes. Repressors, operators, and promoters are binding sites on the DNA where enzymes attach to begin or prevent transcription.

bacterial cells to turn off the *lac* operon when glucose, a preferred food source to lactose, is present in the cell.

V. EUKARYOTIC GENOMES—CHROMOSOMES

The genomes of eukaryotic organisms are characterized by DNA in linear molecules with associated proteins in a complex structure known as chromatin, which is contained within a membrane-bound nucleus. Chromatin is located in a diffused state during interphase of the cell cycle when transcription and translation occur. When the chromatin condenses during cell division and becomes visible, the heavily staining structures are called chromosomes. The number of chromosomes found in different eukaryotic groups vary considerably, but each group displays a characteristic range of chromosome numbers. For example, humans and great apes have haploid chromosome numbers of $N = 23$ and $N = 25$, respectively, whereas different groups of parasitic Hymenoptera (wasps) have $N = 5$ and $N = 10$.

Eukaryotic genes contrast sharply with prokaryotic genes in both structure and regulation of gene expression (Fig. 4). First, whereas several prokaryotic genes are grouped together under control of a single promoter, each eukaryotic gene is expressed individually under control of a promoter that is specific to that gene. Second, eukaryotic genes are often interrupted, with nontranslated introns interspersed between the exons that are translated into proteins. The result is that after mRNA is transcribed in eukaryotes, introns are excised from the mRNA, which is then spliced back together as a mature mRNA for translation into proteins. One consequence of this post-transcriptional modification is that the same gene can be spliced into several different mature mRNAs by the inclusion or exclusion of different exons. A classic example of how alternative splicing controls development in animals is through the cascades of gene expression that determine sex and control sexual differentiation in animals (Hodgkin, 1990; Marin and Baker, 1998). In *Drosophila melanogaster*, for instance, the male and female forms of the gene *Doublesex* both have their first three exons in common, but differ in which 3′ exons are included in the transcript after splicing.

Because most eukaryotes are multicellular, they face three other challenges that do not apply to prokaryotes. First, development from a single-celled egg to an adult requires regulated and coordinated changes in gene expression throughout the life span of an organism. Second, after tissues differentiate, they have unique suites of genes expressed within them. Think of the differences between a heart muscle cell and a secretory cell in the pancreas. Third, cells must communicate with each other, so that much of the control of expression of genes within a cell may come from hormones or other signaling molecules that diffuse or are actively taken up into the cell.

Most regulation of gene expression in eukaryotes is through positive control. A transcription factor must bind with the promoter to initiate transcription of the message by RNA polymerase. However, the ability of the transcription factor to bind to the promoter will be influenced by other proteins that bind to both the enhancer sequences on the DNA and the transcription factor. Enhancer sites are usually sequences of DNA less than 20 nucleotides long. Promoter binding regions will often include a specific sequence recognized by a transcription factor. A common promoter site consists of a TATA nucleotide sequence, thus the name "tata box."

Another common feature of eukaryotes is gene duplication. Some genes that are important for basic housekeeping, like rRNA, may have hundreds or thousands of copies of the gene within the genome, which results in functional redundancy. Another form of duplication is represented by genes coding for blood proteins (globins) in vertebrates. These gene families in-

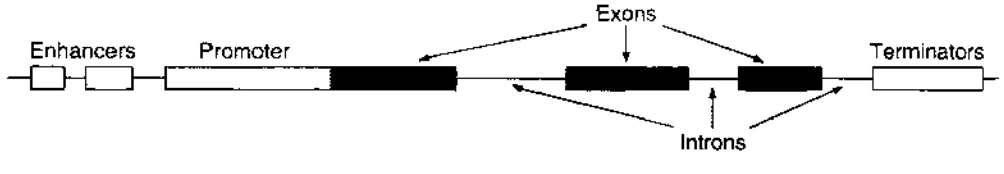


FIGURE 4 General structure of a eukaryotic gene, including the upstream promoter and enhancer sequences where protein transcription regulators can bind and the downstream terminator where transcription ends. Eukaryotic genes are interrupted by exons, which are translated or transcribed, as well as nontranslated introns, which are removed during the process of post-transcriptional modification.

clude several kinds of duplications, where gene clusters code for a diversified set of proteins that are part of fetal blood versus adult blood on the one hand, and a series of nonfunctional pseudogenes on the other hand.

A. Noncoding Repetitive DNA Sequences

A large proportion of the genomes of eukaryotes and some prokaryotes consists of repetitive DNA, which is completely untranscribed and consists of tandemly arranged sequences with identical nucleotide composition that can be repeated up to several million times within the genome. Some repetitive DNA is tied up in the heterochromatic parts of the genome that make up structural elements of chromosomelike centromeres (where sister chromatids bind during cell division) and telomeres (the ends of chromosomes). Otherwise, repetitive DNA may be interspersed throughout the genome. Repetitive elements can be very large (>10,000 nucleotides) or of moderate size (100–800 nucleotides), but the majority of highly repetitive sequences consist of repeats smaller than 100 nucleotides. Generally speaking, repetitive DNA makes up more than half of eukaryotic genomes, and much of the repetitive DNA is now thought to arise as a by-product of transposable elements moving within genomes (Li, 1997).

Repetitive DNAs are highly variable among individuals in a population and are generally classified according to the size of the repeat unit, the number of repeat units per array, and the genomic location of the tandem arrays. There is presently no consensus in the classification of simple tandem repeats and the terminology used varies among authors. The term "microsatellite" denotes any tandem array of repeats with a unit length of 2–5 base pairs. The most common human microsatellites are dinucleotide arrays of $(CA)_n$, which means *n* repeats of CA. There are approximately 50,000 $(CA)_n$ arrays in the human genome or one array for every 30,000 nucleotides. It has been estimated that there are 300,000 trinucleotide and tetranucleotide microsatellites in the human genome or, on average, one array in every 10,000 nucleotides of genomic DNA.

Microsatellites are used extensively in population genetics, DNA fingerprinting for forensics or paternity analysis, and genome mapping. Because a large number of alleles are usually found at each microsatellite locus, a collection of several microsatellite loci can lead to a huge number of possible combinations or DNA profiles, making these loci extremely powerful for identity analysis and forensic use. For example, in the investigation of a crime where a biological specimen has been obtained from the crime scene, DNA from the specimen and from the suspect are typed for size at several unlinked microsatellite loci. If the profiles do not match, the suspect is excluded. If they match, the probability of this match occurring by chance can be calculated from the frequencies of alleles in the population. This probability is usually extremely low because of high polymorphism at microsatellite loci. When primers for microsatellites are well designed, and when each pair of microsatellite primers amplifies only a single locus (e.g., it is possible to have similar microsatellites embedded within repeated fragments of large size, so that a single primer pair amplified two or more loci), the high polymorphism and the ease of typing of microsatellite loci make them very useful for genetic mapping, studies of population structure, detecting loss of heterozygosity, and studies of behavior.

VI. ORGANELLES AND THEIR DNA

The most likely origin of organelles within the cells of eukaryotic organisms is ancient bacteria or blue-green algae taken up as endosymbionts into the cells of early eukaryotes. The strongest line of evidence for an endosymbiotic origin is that organelle genomes most closely resemble those of prokaryotes, both in general characteristics and in the details of DNA sequences within genes. Like prokaryotic genomes, organelle genomes are circular double-stranded molecules with very small amounts of repetitive DNA. In addition, organelle genes do not contain introns, are transcribed polycistronically, and have relatively simple prokaryotelike promoters (Mayfield *et al.*, 1995). Although organelles carry out their own transcription and translation, gene expression and regulation in organelles depend on regulatory proteins that originate from nuclear genes.

A. Mitochondrial DNA in Animals

The animal mitochondrial genome (mtDNA) is a single circular, double-stranded DNA molecule that ranges in size from 14,000 to 42,000 nucleotides. With few exceptions the animal mitochondrial genome contains 37 genes that encode 13 protein subunits, 22 tRNAs, and a small and a large ribosomal subunit RNA. Several compilations of primers that are useful for polymerase chain reaction (PCR) amplification of segments of the mitochondrial genomes of animals are currently available (Avise, 1994; Simon *et al.*, 1994; Palumbi, 1996; Roehrdanz and Degrugellier, 1998).

The analysis of mitochondrial DNA has become one of the most powerful tools for studying animal populations. The mitochondrial genome is primarily maternally inherited as a haploid genome. It does not recombine and it mutates at a faster rate than the nuclear genome in most animal groups. For these reasons mitochondrial DNA can be used to estimate phylogenetic relationships among maternal lineages. Sequence differences arising from mutations in mtDNA haplotypes record the phylogenetic histories of female lineages within and among species. Variation in the frequencies of different mitochondrial haplotypes can be used to estimate effective migration rates among populations and genetic diversity within and among populations. This level of variation is also used for estimating phylogenetic relationships among recently evolved taxa. There are a number of unique perspectives on evolution that have been prompted by genetic findings on animal mitochondrial DNA. These include the usual pattern of within-individual homoplasmy (predominance of a single mtDNA sequence) despite between-individual sequence differences, and the rapid pace of nucleotide substitution.

B. Mitochondrial DNA in Plants

Plant mtDNA is extremely different from animal mtDNA, and this fact has prompted several investigators to suggest independent symbiotic origins. The mtDNA genome in plants ranges from 200,000 to 2,400,000 nucleotides in circumference, and typically exists as a collection of different-sized circles arising from extensive recombinational processes within individuals that convert between a "master" molecule and subgenomic circles. Inheritance is usually (but not always) maternal. Although plant and animal mtDNAs are similar with regard to gene content and general function, their evolutionary patterns are diametrically different. Plant mtDNA appears to evolve rapidly with respect to gene order, but slowly in nucleotide sequence (perhaps 100-fold slower than animal mtDNA). Reasons for the slow accumulation of point mutations are not understood, but plant mtDNA may possess relatively error-free DNA replication systems, or perhaps highly efficient enzymes for repair of DNA damage. The lack of nucleotide-level variation and regular recombination have limited the utility of plant mtDNA in population genetics.

C. Chloroplast DNA (cpDNA)

The cpDNA molecule varies in size from about 120,000 to 247,000 nucleotides in circumference in photosynthetic land plants, due largely to reiteration of a large inverted repeat that includes genes for the ribosomal RNA subunits. There are from 22 to 900 copies of the genome in each chloroplast and each encodes 123 genes. These include 4 genes that encode rRNA, 20 genes for the ribosomal proteins, 30 genes for the tRNAs, many proteins that are important for photosynthesis, 6 of the 9 genes for the synthesis of ATPase, and chloroplast RNA polymerase. Unlike mitochondrial DNA, 15 of the cpDNA genes contain introns. Most chloroplast proteins are encoded by the nucleus. For example, chloroplast ribosomes consist of 52 proteins, only 19 of which are encoded by the plastid genome.

Chloroplast DNA is transmitted maternally in most flowering plants, biparentally in a few, and *paternally* in gymnosperms. These cpDNA genes have been shown to transpose to the nucleus and there is good evidence that mtDNA, cpDNA, and nuclear genomes exchange genes. The rate of cpDNA evolution generally appears to be slow in terms of both primary nucleotide sequence and gene rearrangement. Because of the large size of the cpDNA genome, most systematic treatments have involved restriction site or sequence determinations for particular genes or have monitored the taxonomic distributions of unique cpDNA structural features across higher-level plant taxa. Nonetheless, some studies have uncovered considerable intraspecific cpDNA variation as well, and this suggests that portions of the cpDNA should be useful in population genetics. Several primers for amplifying cpDNA genes in a number of plant species are available (e.g., ndhF: Olmstead and Sweere, 1994; rbcL: Palumbi, 1996).

VII. MUTATION

Mutation is the ultimate source of genetic variation and is generally defined as any change in genetic material, whether phenotypic effects caused by the mutation are large, small, or nondetectable. The causes of mutation include direct insults from environmental effects like ultraviolet light, chemical solvents, X rays, and other atomic particles, but many mutations arise from errors in DNA replication and DNA repair during cell division.

One class of mutations arises from gross changes at the level of chromosomes, where chromosomes break and are then modified by deletions, insertions, inversions, translocations, and duplications (see Table II). Although the term "chromosome" is used here, similar mutations affect the genomes of prokaryotes as well. Chromosomal mutations can result in changes in the total amount of DNA found in different species and, in the case of gene duplications or deletions, mutations can change the number of genes. Mutations like these can have drastic phenotypic effects. Individual genes can be disrupted if insertions occur within them, or if parts of a gene are lost in a deletion. Transposable elements cause deleterious mutations by inserting themselves into genes.

Inversions of chromosomal fragments, where a fragment is flipped 180° from its original orientation and reinserted at the same place, will cause ballooning of the chromosomes when they pair during meiosis and will prevent recombination. Inversions are common in some organisms, like the Diptera (especially *Drosophila*). The lack of crossing-over and recombination means that genes within an inversion will be inherited intact as a "supergene."

Translocation of pieces of chromosomes from one place to another within the genome will result in change in the size or number of chromosomes. The effects of translocations can be dramatic, often resulting in sterility because of the inability of the altered chromosomes to pair during meiosis. Thus, translocations are a potential mechanism of isolation of a single species into incompatible breeding groups, leading eventually to the formation of new species.

The magnitude of phenotypic effects of duplications depends on the size of the affected chromosomal fragment, and thus the number of genes that are duplicated. An example is trisomy 21, which results in Down syndrome (mental retardation, flat facial features, heart defects). Trisomy results from a failure of proper segregation of chromosomes during meiosis so that one gamete contributes two complete copies of chromosome 21, resulting in three copies in the zygote. On the other hand, duplications are an integral part of the evolution of multigene families, like the various globin genes and nonfunctional pseudogenes in vertebrates. Duplication is the main mechanism by which entirely new genes arise (Li, 1997).

Mutations can also occur at the level of single-nucleotide changes, including some that affect the reading of the genetic code (see Table II). Single-nucleotide substitutions are classified as transitions if they occur between similar nucleotides, that is, between purines (A ↓ G) or pyrimidines (C ↓ T), and as transversions if mutations substitute purines for pyrimidines and vice versa (A, G ↓ T, C). This distinction is important because transitions are much more common than transversions, sometimes occurring 20 times more frequently. Further examination of the genetic code (see Table I) reveals that where a substitution occurs within a codon will determine the severity of the mutation. Most mutations in first nucleotide position will result in an amino acid change (mis-sense) or a mutation to a stop codon (non-sense). Mutations at the second position always result in a change (mis-sense and non-sense). Third-position mutations are usually synonymous ("silent"), resulting in no change of amino acid. Nucleotide insertions and deletions will result in a shift of the "reading frame," so that all amino acids after the mutation are changed.

Mutations occur at very low rates, and are usually measured as the number per meiosis in eukaryotes, or as the number per cell division for prokaryotes. The numbers of new mutations are commonly less than 1 in 10,000 (10^{-4}) for nucleotide substitutions and are as low as 1 in 100,000,000 (10^{-8}) for some visible mutations like coat colors. In an evolutionary sense, mutation is considered a very slow process. However, mutation still can produce considerable variation within populations because the numbers of new mutations produced in each generation in a population will be a function of population size. In addition, if the number of genes in a genome is taken into account, even low mutation rates can generate high numbers of new mutations in each individual within a population.

See Also the Following Articles

ADAPTATION • BACTERIAL GENETICS • DIVERSITY, MOLECULAR LEVEL • ECOLOGICAL GENETICS • SPECIATION, PROCESS OF

Bibliography

Avise, J. C. (1994). *Molecular Markers, Natural History, and Evolution.* Chapman & Hall, New York.

Gutell, R. R. (1996). Comparative sequence analysis and the structure of 16S and 23S rRNA. In *Ribosomal RNA: Structure, Evolution, Processing, and Function in Protein Biosynthesis* (R. A. Zimmerman and A. E. Dahlberg, eds.). CRC Press, Boca Raton, Florida.

Hartl, D. L., and Jones, E. W. (1998). *Genetics: Principles and Analysis,* 4th ed. Jones and Bartlett Publishers, Sudbury, Massachusetts.

Hodgkin, J. (1990). Sex determination compared in *Drosophila* and *Caenorhabditis. Nature* 344, 721–728.

Lewin, B. (1997). *Genes VI.* Oxford University Press, Oxford, United Kingdom.

Li, W.-H. (1997). *Molecular Evolution.* Sinauer Associates, Sunderland, Massachusetts.

Marin, I., and Baker, B. S. (1998). The evolutionary dynamics of sex determination. *Science* 281, 1990–1994.

Mayfield, S. P., Yohn, C. B., Cohen, A., and Damon, A. (1995). Regulation of chloroplast gene expression. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* 46, 147–166.

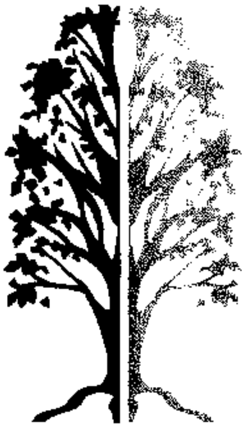
Olmstead, R. G., and Sweere, J. A. (1994). Combining data in phylogenetic systematics: An empirical approach using three molecular data sets in the Solanaceae. *Systematic Biol.* 43, 467–481.

Palumbi, S. R. (1996). Nucleic acids II: The polymerase chain reaction. In *Molecular Systematics* (D. M. Hillis C. Moritz and B. K. Mable eds.), 2nd ed. Sinauer Associates, Sunderland, Massachusetts.

Prusiner, S. B., and Scott, M. R. (1997). Genetics of prions. *Annu. Rev. Genetics* 31, 139–175.

Roehrdanz, R. L., and Degrugillier, M. E. (1998). Long sections of mitochondrial DNA amplified from fourteen orders of insects using conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Amer.* 91, 771–778.

Simon, C., Frati, F. A., Beckenbach, L., Crespi, B., Liu, H., and P. Flook. (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Amer.* 87, 651–701.



GENETIC DIVERSITY

Eviatar Nevo
University of Haifa

GLOSSARY

electrophoresis Separating charged molecules (such as polypeptides or polynucleotides) between the two poles of an electric field.

genome A complete single set of genes of an organism or organelle; also the basic haploid chromosome set.

genomics The study of the molecular organization of genomic DNA and physical mapping.

heterozygosity The average number of different heterozygotes across loci divided by all loci studied.

polymorphism Different alleles at a gene locus within a population.

GENETIC DIVERSITY (i.e., molecular hereditary differences within or between populations) is the basis of evolutionary change. The nature of genetic diversity among organisms has always been the basic problem of evolutionary genetics (Darwin, 1859; Lewontin, 1974; Kimura, 1983) as well as of domestication, agriculture, and medicine. However, despite its cardinal role in evolutionary theory and application, the maintenance of genetic diversity remains largely enigmatic, notwithstanding the dramatic discoveries of molecular biology, which revealed abundant genetic diversity in nature.

I. THE PROBLEM

A. Historical Background

The era of molecular evolution was ushered in by three major discoveries in the late 1950s and 1960s: (i) recognition by Markert and Moller in 1959 of molecular diversity of enzymes and their importance in genetics, physiology, development, and evolution; (ii) Identification by Zuckerkandl and Pauling in 1965 of protein sequence variation between species and the resulting postulate of the molecular clock; and (iii) evaluation in 1966 of enzyme variation (isozymes and allozymes) in *Drosophila* by Lewontin and Hubby and Johnson *et al.* and in humans by Harris. Molecular biology permitted the characterization of genetic diversity among individuals, populations, and species—the three cornerstones of evolution. It did so first by unraveling relationships between genes and proteins (Lewontin, 1974). Second, it did so by elucidating, at the extranuclear and nuclear coding and noncoding DNA regions, the structure, expression, function, mechanism, and evolution of genes, intergenic spacers, and multigene families by employing recombinant DNA methodologies (Watson *et al.*, 1987; Avise, 1994).

Advancements in molecular evolution have followed

Encyclopedia of Biodiversity, Volume 3

the introduction of new laboratory techniques (Avise, 1994). Among the most influential methods have been protein electrophoresis in the late 1960s and 1970s, restriction fragment length polymorphism (RFLP) analyses of nuclear and mtDNA in the late 1970s and 1980s, DNA fingerprinting in the mid to late 1980s, and polymerase chain reaction (PCR)-mediated DNA sequencing in the 1990s. Comparative genomics and genome prospecting (*Science* 286, October 15, 1999) promise to highlight the genetic basis of evolutionary change and the nature, expression, meaning, transfer, and regulation of information in biological systems, thus unraveling the blueprint and evolutionary forces driving life.

B. The Enigma of the Maintenance of Genetic Diversity

Early and later molecular–genetic studies revealed massive levels of genomic diversity, surpassing any prediction. In addition, a comprehensive compilation of dense genetic maps was published of viruses, bacteria, plants, animals, and humans. In 1998, Deloukas *et al.* physically mapped 30,000 encoded human genes for proteins of known functions. The ongoing human genome project (HGP), the largest biological project ever (Bodmer and McKie, 1994), ushered in a new era in the life sciences in which complete genomes of prokaryotes (Casjens, 1998) and eukaryotes became available. Genetic maps, current complete genome sequences and those expected in the near future, and tens to hundred of thousands of protein and DNA sequences have contributed dramatically to elucidating genome diversity, organization, expression, dynamics, and evolution (Karlin *et al.*, 1998). They form the basis for future studies of genetic diversity, when more genomes within species will be analyzed.

However, despite the ease of measuring and deciphering genetic diversity and the dramatic advances in comparative genomics, the evolutionary forces that generate and maintain segregating genetic diversity, preventing allele fixation or random elimination in nature, remain elusive.

C. Theories of Molecular Evolution: Selection versus Neutrality

Theoretically, balancing selection could account for protein polymorphism (Gillespie, 1991). In contrast, the neutral theory of molecular evolution (Kimura, 1983) suggests that most of the molecular–genetic diversity within and between species is neutral (i.e., nonselective) or "non-Darwinian." The neutralist–selectionist debate has been one of the major controversies in evolutionary biology since the late 1960s. How much of the genetic diversity at single and multilocus structures is adaptive, processed by natural selection and contributing to differences in fitness? The problem of distinguishing between deterministic and stochastic forces in evolution has pervaded evolutionary biology at all levels, genotypic and phenotypic, and is now focused on DNA polymorphisms. I recognize the contribution of the neutral and nearly neutral theories of molecular evolution, primarily by representing a null hypothesis to selection. Nevertheless, by ignoring the ecological heterogeneity and stress in evolution, neutral and nearly neutral theories have stripped genetic diversity from nature. I believe that in-depth understanding of genetic diversity in nature is intimately linked to the interface between ecology and genetics; hence, to ecological genetics and now to ecological genomics. I submit that only this essential interface can meaningfully highlight the dynamic evolution of genetic diversity in nature.

D. The Structure of This Review

This review comprises three unequal perspectives of molecular–genetic diversity of protein and DNA in nature: methodology, evidence, and theory. The main focus is on the huge amount of protein and DNA evidence from nature in diverse groups of organisms across phylogeny from bacteria to humans and across diminishing geographical scales (global, regional, and local). The evidence naturally leans heavily on results of genetic diversity in plants and animals obtained during the past 25 years at the Institute of Evolution, University of Haifa, Israel, that demonstrate the pervasiveness of ecological determinants in genetic differentiation. In theory too, I focus on theoretical results obtained during the past decade at the Institute of Evolution that show how selection in cyclical environments can maintain genetic polymorphism in nature, thus preventing drift. I mainly refer to books which cite the primary literature.

II. METHODOLOGIES

The methodologies analyzing protein and DNA diversity have been extensively described (Lewontin, 1974; Avise, 1994; Mitton, 1997), so they are only briefly reviewed here.

A. Protein Polymorphism

1. Technical Innovation

New techniques at the protein level permit an increasingly subtle resolution of allelic variation at a locus. These techniques include isoelectric focusing in various media; high-resolution, two-dimensional electrophoresis; thermal and urea denaturation analysis; and sequential gel electrophoresis (SAGE), in which electromorphic classes were retested with other pHs, buffer systems, and gel pore sizes (Ramshaw *et al.*, 1979).

2. Hidden Variation

"Alleles" detected by routine gel electrophoresis are essentially phenotypes or "electromorphs" (i.e., internally heterogeneous, genetically involving "hidden variation"). For example, xanthine dehydrogenase has many more alleles than are visible routinely (Ramshaw *et al.*, 1979). Recent pulsed-field gel electrophoresis for separating large DNA fragments, recombinant DNA techniques, methods of genomic "hopping," and new DNA sequencing strategies and genomic analysis reinforce the powerful field of molecular evolution, climaxing in the HGP (Bodmer and McKie, 1994). Extensive comparative analyses across genomes of model organisms (Karlin *et al.*, 1998) became the focus of genomic and proteomic studies, highlighting diversity in nature (Avise, 1994). These new horizons partly unraveled the molecular structure, function, and evolution of life. Bioinformatics became increasingly important as many genes and genomes became analyzable. Comparative genomics, using dense genetic maps based on coding genes and on microsatellite and single nucleotide polymorphisms (SNPs), permit precise gene homolog alignment across taxa, the unraveling of gene function and regulation, highlighting of genome organization and evolution, and the determination of the genetic basis of speciation and adaptation.

B. DNA Polymorphism

1. RFLP, PCR, and Mini- and Microsatellites

Many assays reveal DNA diversity (Avise, 1994). Before the revolutionary PCR was established, DNA fingerprinting technology had been revolutionized by RFLP analysis, a very efficient technique but quite laborious and not suitable for high-throughput applications. Hundreds of restriction enzymes, originating in bacteria, cleave duplex DNA at particular oligonucleotide sequences, usually 4–6 base pairs (bp) in length. Since its development, PCR has been a powerful tool for DNA fingerprinting and for effectively measuring genetic diversity within and between populations. These molecular markers include minisatellite, also called variable number of tandem repeats (VNTR), consisting of core sequences 10- to 15-bp long and extending from 16 to 64 bp, and microsatellites (2–6 bp), also called simple sequence repeats (SSRs) (Goldstein and Schlotterer, 1999). Microsatellites and minisatellites differ in size, mutation processes, and chromosomal distribution, but the boundary between the two classes is not defined.

2. Random Polymorphic DNA

A novel PCR-based strategy involving the use of arbitrary primers (AP-PCR) to amplify random polymorphic DNA (RAPD) fragments has been developed using short primers (≈10 bp). RAPDs generate polymorphic markers using very small amounts of starting DNA or RNA, independently of any prior knowledge of the target DNA sequence. This feature makes RAPD a useful tool in genetic analysis. Compared with protein polymorphisms, DNA markers reveal more genetic diversity and permit choice among sets of loci with different functions and patterns of inheritance (Mitton, 1997; Avise, 1994). Use of molecular markers dramatically widened the perspectives of evolutionary biology (Avise, 1994).

3. Amplified Fragment Length Polymorphism: A New, Powerful Technique for DNA Fingerprinting

A major novel DNA fingerprinting technique called amplified fragment length polymorphism (AFLP) was recently described. AFLP is based on the selective PCR amplification of restriction fragments from a total digest of genomic DNA. AFLP can analyze genetic diversity in many hundreds of genes in both coding and noncoding regions across the genome if several enzyme combinations are applied, allowing high-resolution genotyping of fingerprinting quality. The time and cost efficiency, replicability, and resolution of AFLPs are superior or equal to those of other markers (allozymes, RAPD, RFLP, and microsatellites), except that AFLP methods primarily generate dominant rather than codominant markers.

4. Nucleotide Polymorphism

Studies of nucleotide polymorphism have three distinct advantages: (i) detection of all types of sequence changes (single nucleotide substitutions, insertion/deletions, and copy number variation in nucleotide repeat motifs), of which SNPs are the most common; (ii) variant detection in coding (cSNPs and noncoding DNA; and (iii) combined detection and genotyping using a single method.

5. Summary

DNA techniques revealing DNA diversity include RFLP, RAPD and AFLP, mini- and microsatellites, SNP and sequencing; all are relevant for analyzing DNA diversity. PCR is currently the method of choice for amplifying DNA segments for detecting polymorphism. Generally, DNA diversity is higher than protein diversity (see Fig. 4).

C. Comparative DNA Analysis across Diverse Genomes

The era of comparative genomics is dramatically advancing (Casjens, 1998; Karlin *et al.*, 1998; *Science*, October 15, 1999), including the analysis of complete genomes for various organisms and imminent completion of the HGP. Genes, individuals, and species are becoming comparable. Sequence polymorphism will highlight the generation, maintenance, and function of genetic diversity responsible for controlling normal development, physiologic homeostasis, and disease processes. This information explosion will revolutionize molecular evolutionary studies. Comparative analysis of complete genomes includes assessments of genomic compositional contrasts based on di-, tri-, and tetranucleotides relative abundance values; identification of rare and frequent oligonucleotides; evaluations and interpretations of codon biases in several large prokaryotic genomes; and characterizations of compositional asymmetry between the two DNA strands in certain bacterial genomes. Comparative analysis also allows identification of alien (e.g., laterally transferred) genes and detection of potential specialization islands in bacterial genomes and the assessment of DNA curvature. It can compare genomes within species and across life and cluster organisms according to linguistics based on many words in large DNA stretches.

D. Microarrays: Biotechnology's Discovery Platform for Functional Genomics

Advances in microarray technology make possible massive parallel mining of biological data, based on PCR, with biological chips providing hybridization-based expression, monitoring, polymorphism detection, and genotyping on a genomic scale, as reported by Schena *et al.* (1998). Microarrays containing sequence representative of all the genes of an organism may soon permit the expression analysis of the entire human genome in a single reaction. These "genome chips" will provide unprecedented access to genomic diversity of many thousands of genes to large-scale gene discovery as well as polymorphism screening and mapping of genomic DNA clones on a massive scale—critical for science and application in agriculture and medicine. Oligonucleotide microarray (DNA chip)-based hybridization analysis is a promising new technology which potentially allows rapid and cost-effective screens for all possible mutations and sequence variations in genomic DNA. Currently, it is performed in humans, mice, and *Arabidopsis*, but use in additional species has been reported. The increasing use of relatively inexpensive microarrays is expected to revolutionize genomic, proteomic, and other biological research projects.

E. Encyclopedia of Genes

Extensive ongoing programs in plants and animals are generating a large database of expressed sequence tags (ESTs) that can provide rapid access to numerous genes and their diversity. The development of a comprehensive database of ESTs for *Arabidopsis*, corn, soybean, rice, barley, wheat, mouse, and human has been reported. As of October 23, 1998, 352,040 sequences had been generated in the mouse and annotated and deposited in dbEST, in which they comprised, according to Marra *et al.* (1999), 93% of the total ESTs available for the mouse. EST data are versatile and have been applied to gene identification, comparative sequence analysis, comparative gene mapping and candidate disease gene identification, genome sequence annotation, microarray development, and the development of gene-based map resources. Large-scale exploration of the genic diversity of the genome is becoming possible through the UniGene database (*http://www.ncbi.nlm.nih.gov/UnGene*).

III. EVIDENCE: GENETIC DIVERSITY WITHIN AND AMONG SPECIES

A. Protein Diversity

Electrophoretic results have revealed large amounts of genic polymorphism in natural populations (Lewontin, 1974; Nevo, 1978, 1988, 1998; Hamrick *et al.*, 1979; Nevo *et al.*, 1984; Avise, 1994; Mitton, 1997). This is true for both eukaryotes and prokaryotes. Currently, theory appears to lag behind evidence. The review and reanalysis of approximately 1100 plant and animal species (Nevo *et al.*, 1984) revealed an average heterozygosity, *H*, of 0.073 (SD 0.076) and an average proportion of loci polymorphic, P, of 0.284 (SD 0.197). Across species, the coefficient of correlation between *H* and *P* was $r = 0.793$, $p < 0.001$. These estimates are based on the average of 23 electromorphically detectable gene loci (minimum 14 loci) and an average of 199 individuals per species (minimum 10 individuals) (Fig. 1). This set of data was assembled up to 1983, and many hundreds of species have since been tested (see Fig. 2.2 in Avise, 1994, that summarizes heterozygosity estimates of 1803 species: 648 vertebrate, 370 invertebrates, and 785 plant species), all demonstrating diverse levels of polymorphisms (Mitton, 1997). Genetic diversity varies dramatically among species. Fruit flies, marine mussels, and conifers have much genetic diversity, whereas large vertebrates have much less (Mitton, 1997).

The aforementioned genetic indices appear to be lower than the real ones since routine horizontal gel electrophoresis underestimates diversity and protein coding genes represent only a small portion of the genome. Regarding levels of diversity, protein diversity does not represent the noncoding genome, which amounts to 90–98%, involving invariant and variant portions whose general diversity levels in natural populations appear to be high. Studies of "hidden variation" indicate that most protein diversity is detectable (Ramshaw *et al.*, 1979), thus increasing *H* of enzymatic loci. Moreover, at the DNA sequence level of individuals, most loci may prove heterozygous. Nevertheless, there are significant merits to the estimates of *H* and *P*, derived from routine protein electrophoresis. First, although relative, low, and limited genomically, they are commensurate estimates of genetic diversity among numerous populations and species living in varied ecologies. Second, *H* and *P* are highly correlated in global (Nevo *et al.*, 1984) and regional (Nevo, 1988, 1998) analyses. Thus, although hidden variation is regrettably lacking in studies of routine protein electrophoresis, in most species the high correlation of *H* and *P* makes for a profitable analysis since *P* is largely independent of the high resolution achievable by SAGE and other new techniques. Third, the relatively large number of electromorphic protein loci currently available in *Drosophila* and the many species analyzed permit a sound statistical analysis of isozyme phenotypes in an ecological context rather than only in genetic terms (Nevo *et al.*, 1984; Nevo, 1978, 1988, 1998).

1. The Adaptive Evolution of Enzyme Kinetic Diversity

Kinetic studies of hemoglobin, haptoglobin, and transferin proteins and at least a dozen enzyme polymorphisms typically reveal biochemical kinetic differences among the gene products of alternative genotypes at a locus (Mitton, 1997). These include single gene effects such as lactate dehydrogenase in killifish, leucine aminopeptidase in blue mussel, and phosphoglucose isomerase (PGI) in *Collias* butterflies. In the latter, Watt

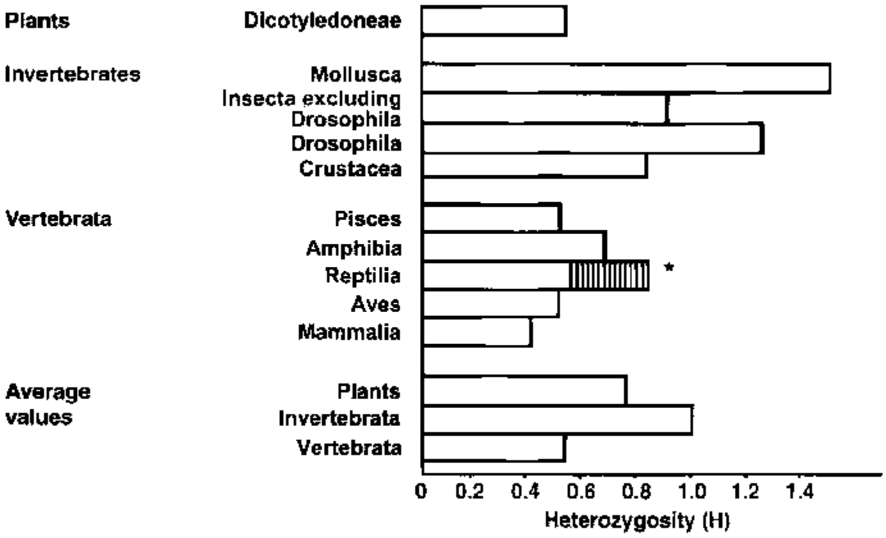


FIGURE 1 Levels of genetic (allozymic) diversity among higher taxa based on a mean of 23 gene loci per species, in 1111 species. *Hatched region indicates parthenogenetic species (reproduced with permission from Nevo *et al.*, 1984).

and colleagues discovered in the 80s that PGI heterozygotes fly over a greater range of temperatures and produce 33% more eggs. Additional examples are alcohol dehydrogenase (ADH) in fruit flies, salamanders, and barley; glutamate pyruvate transaminase in copepods; and esterase, glucose-6-phosphate, 6-phosphogluconate dehydrogenase, and superoxide dismutase in fruit flies. In all these cases, biochemical–kinetic studies reveal differences among phenotypes, either within or between species, that have measurable effects of alternative genotypes on the physiology of whole individuals. They result in fitness differences distinguished by natural selection in accord with their alternative spatiotemporal environments (Mitton, 1997). Usually, the heterozygotes are intermediate between homozygotes, but sometimes molecular overdominance for enzyme kinetics is found.

2. Patterns of Variation among Loci

Proteins vary genetically. Regulatory enzymes appear to be more variable than nonregulatory enzymes, and enzymes that work on many substrates appear to be more variable than enzymes that utilize a single substrate. Genetic diversity tends to decrease from monomer to dimer to tetramer as the steric restrictions on the molecule increase. The number of alleles at a locus and the heterozygosity at a locus tend to increase with the subunit size of the protein.

3. Genetic Population Structure

Geographical clines may reflect the action of natural selection on genetic polymorphism at local, regional, and global scales. In *Drosophila melanogaster*, several latitudinal clines occur for many characters such as allozymes, inversions, and quantitative traits. The identical nature of these clines on the various continents, in both the Northern and Southern Hemispheres, strongly suggests adaptation to specific stress factors, primarily climatic selection. Polymorphism for stress-resistance genes abounds in natural populations. The ADH polymorphism shows high frequencies of the S allele in tropical regions and this declines with latitude. The reasons for this cline are difficult to determine because of the entanglement with other polymorphisms varying with latitude. In 1977, Van Delden and Kamping reviewed the tentative connections with other polymorphisms such as α-GPDH, in(2L)t inversion, body size, and development time with respect to the possible environmental stress factors involved. They concluded from these results, and also from recent experiments, that the (2L)t plays a dominant role in resistance to high temperature and is partly responsible for the ADH cline. They are currently studying the specific selective forces acting on ADH, focusing on the physiological and life history aspects. Many plant and animal species distributed in several climatic zones—tropical, temperate, and arctic—display a decline in gene diversity, both allozyme and DNA, towards the Arctic (Nevo *et al.*, 1984). This is highlighted by postglacial colonizing populations. Regional and local clines also abound in nature.

4. Partition of Genetic Diversity

An important analysis is that of genetic partition within and between populations. In most analyzed outcrossers, genetic diversity is chiefly within populations (e.g., 85% in humans). In contrast, in some selfers (e.g., wild emmer wheat) only 40% of genetic diversity is within and 60% is between populations. This partition is shaped by both external (gene flow and selection at the population level) and internal (genomic and genic) evolutionary forces.

B. DNA Diversity

DNA technologies revolutionized population genetics by providing an unprecedented amount of genetic diversity for critical analysis and hypothesis testing. I briefly review these dramatic developments, from molecular (extranuclear, mitochondrial, chloroplast, and nuclear) DNA markers to sequence polymorphism.

1. MtDNA Diversity

MtDNA is largely maternally nonrecombiningly inherited and demonstrates a rapid pace of evolution and an extensive high level of intraspecific restriction. PCR and sequence polymorphism derive primarily from base substitution with a preponderance of transitions, length variation, sequence rearrangements, and duplications of coding sequences (Avise, 1994). Hence, its relevance is not only to large-scale phylogenies but also to microgeographic divergence, i.e., to intraspecific phylogeography, the recently established bridge between population genetics and systematics. The amount of mtDNA polymorphism within species and sequence heterogeneity is striking. Remarkably, a different scenario emerged for plant mtDNA from that in animals. Whereas animal mtDNA ranges in size from 14 to 26 kb in length, that of plants (e.g., maize) is approximately 30 times larger. Most polymorphism in plant mtDNA is attributable to major reorganizations of sequence. The level of mtDNA polymorphism in higher animals is several-fold higher than that of single-copy nuclear DNA. The maternal inheritance and high polymorphism of mtDNA and the assay of whole mtDNA by long PCR provide unique opportunities for population evolutionary studies of animals and plants on both micro- and macrogeographic scales. They contribute to natural history, population structure and signature, gene flow, hybridization, biogeography, phylogeny, and biological conservation.

Mitochondrial genomes are increasingly being used to study ancient divergences among animal groups. Recent studies by Curde and Koder (1999) of complete mitochondrial DNA sequences and reached somewhat heretical conclusions, raising questions about the use of mitochondrial gene sequences for studying the relationships among highly divergent lineages. Other studies have documented convergent evolution of mitochondrial gene order, casting doubt on the use of these characters for phylogenetic analysis. The use of mitochondrial genomes for studying such deep divergences is coming under increased scrutiny, and these novel results need to be confirmed with data from nuclear genes.

2. Comparative Summary of Genetic Distances in the Vertebrates from the Mitochondrial Cytochrome b Gene

Mitochondrial cytochrome b is among the most extensively sequenced genes to date across the vertebrates. Johns and Avise (1998) employed approximately 2000 cytochrome b gene sequences from GenBank to calculate and compare levels of genetic distance between sister species, congeneric species, and confamilial genera within and across the major vertebrate taxonomic classes. The results of these analyses parallel and reinforce some of the principal trends in genetic distance estimates derived from multilocus allozymes. In particular, surveyed avian taxa on average show significantly less genetic divergence than do same-rank taxa surveyed in other vertebrate groups, notably amphibians and reptiles.

3. Chloroplastid Genome Comparisons

Chloroplast genome (cpDNA) evolution is also very dynamic, indicating gene transfer to the nucleus. Among the 210 different protein-coding genes contained in the completely sequenced chloroplast genomes from a glaucocystophyte, a rhodophyte, a diatom, an euglenophyte, and five land plants, Martin and colleagues identified in 1998 the set of 45 genes common to each and to a cyanobacterial outgroup genome. Phylogenetic influence, with an alignment of 11,039 amino acid positions per genome, surprisingly indicates that independent parallel gene losses in multiple lineages outnumber phylogenetically unique losses by more than four to one. They identified homologs of 44 different plastid-encoded proteins as functional nuclear genes of chloroplast origin, providing evidence for endosymbiotic gene transfer to the nucleus in plants. cpDNA diversity in lodgepole pine was high within populations, with little (>5%) differentiation among populations, in contrast to the mtDNA pattern as shown by Hamrick and Godt in 1990 (as cited in Mitton, 1997, p. 71). Extensive intraspecific cpDNA diversity may sometime exceed interspecific diversity, reflecting dynamic ecologies and seriously affecting phylogenetic conclusions, as in the pine *Draba* species.

4. Simple Sequence Repeats (SSRs)

SSRs consist of tandem repeats of relatively short nucleotide motifs, such as TCCTCCTCCTCC. Microsatellite repeat number can range from two $(TA)_2$ or three $(GA)_3$ to a few dozen $(GCAA)_{11}$, whereas minisatellites often consist of many dozens or even hundreds of repeated motifs. Significantly, SSRs experience mutation at notably higher rates than do nonrepetitive sequences: 10^{-2} to 10^{-3} per locus, per gamete, per generation, which leads to their high polymorphism. Replication slippage, sister chromatid exchange, unequal crossing-over, and gene conversion may cause microsatellite diversity. Replication slippage seems to play a major role in producing new alleles at microsatellite loci (Goldstein and Schlotterer, 1999). SSRs are characterized by high site-specific and reversible rates of gain or loss in the number of tandem repetitions of a short DNA motif. Importantly, recent critical population genetics and specific chromosome data derived from Li and colleagues in 1999 at our laboratory indicate massive genomic nonrandom chromosomal and environmental distributions of microsatellites, suggesting that they are subjected to natural selection, generating adaptive complexes in contrast to neutral theory expectations.

The functional properties of SSRs suggest evolution's effect on mutability, as recently concluded by King and Soller (as cited in Wasser, 1999). Many SSRs are functionally integrated into the genome and exert a quantitative regulatory effect on gene transcription activity affecting phenotype and fitness. Genes associated with SSRs may be favored by indirect selection whenever quantitative variation in the affected traits can provide a population with genetic resilience for adaptation, especially in stressful, fluctuating, or heterogeneous environments. Such "adjustable genes" may provide a prolific and evolutionarily significant source of quantitative genetic variation.

Nucleotide diversity of the human nuclear genome has been estimated to be approximately 0.1%. For two randomly selected sequences, this number translates into one polymorphic site for 1000 nucleotides (or, in a large sequence sample, one polymorphic site is expected for every 200–500). Since the human nuclear genome contains approximately 3 billion nucleotides, several million polymorphic sites are expected to exist.

5. Single Nucleotide Polymorphism

SNPs are the most frequently found DNA sequence variations in animal and plant genomes, usually followed by SSRs, RAPDs, and allozymes. Recent SNP surveys in humans reported different rates of polymorphism among classes of sites within genes (noncoding, degenerate, and nondegenerate) as well as between genes. Of all coding SNPs, 54% lead to predicted change in the protein sequence. As expected, the coding SNPs that alter amino acid sequence of the encoded protein are found at a lower rate and with lower allele frequencies than silent substitutions. This was interpreted as a reflection of selection against deleterious alleles during human evolution. Determination of ancestral alleles from human SNP polymorphisms became available using high-density oligonucleotide arrays. A densely packed map of human SNP sites could efficiently identify disease-associated genes by linkage disequilibrium between sets of adjacent markers and highlight human history. In plants, SNP polymorphism has been associated with transcript efficiency, nonconcensus splice sites, and gene expression. The following sections discuss the results of genetic diversity and divergence at global, regional, and local scales based primarily on results derived at the Institute of Evolution since 1979.

C. Global Analysis of Genetic Polymorphisms

1. Global Allozyme Diversity across Phylogeny

We used the entire globe as a large-scale ecological genetic laboratory. We analyzed the correlates of biotic factors involving ecological, demographic, and life history variables with the level of genetic diversity in natural populations of animals and plants (Nevo *et al.*, 1984). This review involved 1111 species studied for allozymic variation with an average of 23 gene loci in each species and a biotic profile characterized by 21 variables (7 ecological, 5 demographic, and 9 life history and other biological characteristics). We then (i) estimated the levels of genetic diversity, indexed by heterozygosity and polymorphism, for all species, three major taxa (vertebrates, invertebrates, and plants), 10 different higher taxa, and the categorized 21 biotic factors; (ii) correlated the levels of genetic diversity with the biotic factors; and (iii) matched some of the evidence obtained with theoretical predictions.

The following results were obtained:

1. The levels of genetic diversity vary nonrandomly among populations, species, and higher taxa and also among ecological parameters (life zone, geographical range, habitat type and range, and climatic region) (Fig. 2), demographic parameters (species size and population structure, gene flow, and sociality), and life history characteristics (longevity, generation length, fecundity, origin, and parameters related to the mating system and mode of reproduction).
2. Generally, genetic diversity is higher (i) in species living in broader environmental spectra; (ii) in large species with a patchy population structure and limited migration as well as in solitary or social species; (iii) in species with small body size, annuals or long-lived perennials, older in time, with smaller diploid chromosome numbers; and (iv) in plant species primarily outcrossed, reproducing sexually, and pollinated by wind. Diversity among species has also been summarized by Mitton (1997). Higher genetic diversity characterizes species with high fecundity, large populations, broad geographic ranges, and a high speciation rate. In contrast, low genetic diversity characterizes endemic species with low fecundity and low speciation rate.
3. Genetic diversity is partly correlated and predictable by a three- or four-variable combination of ecological, demographic, and life history variables, largely in that order. Ecological factors account for the highest proportion of the 20% explained genetic variance of all species compared with demographic and life history factors (90, 39, and 3.5%, respectively). Within individual higher taxa, the explained portion of genetic diversity increases considerably (mean of 44% and maximum of 74% in mollusks). Neutrality theory would be substantiated if demographic variables, rather than ecological variables, could better explain the variation in genetic diversity. The opposite, however, was often found, substantiating selection as an evolutionary driving force of genetic diversity in nature. A more detailed global analysis of genetic differentiation in amphibians was performed in 1991 by Nevo and Beiles and in small mammals across the globe (Nevo, 1999). In both cases, genetic diversity, and in amphibians also genome size, is predictable by ecological factors: The higher the environmental heterogeneity and stress, the higher the heterozygosity. Similar conclusions were drawn for 480 species of plants in 1990 by Hamrick and Godt and for

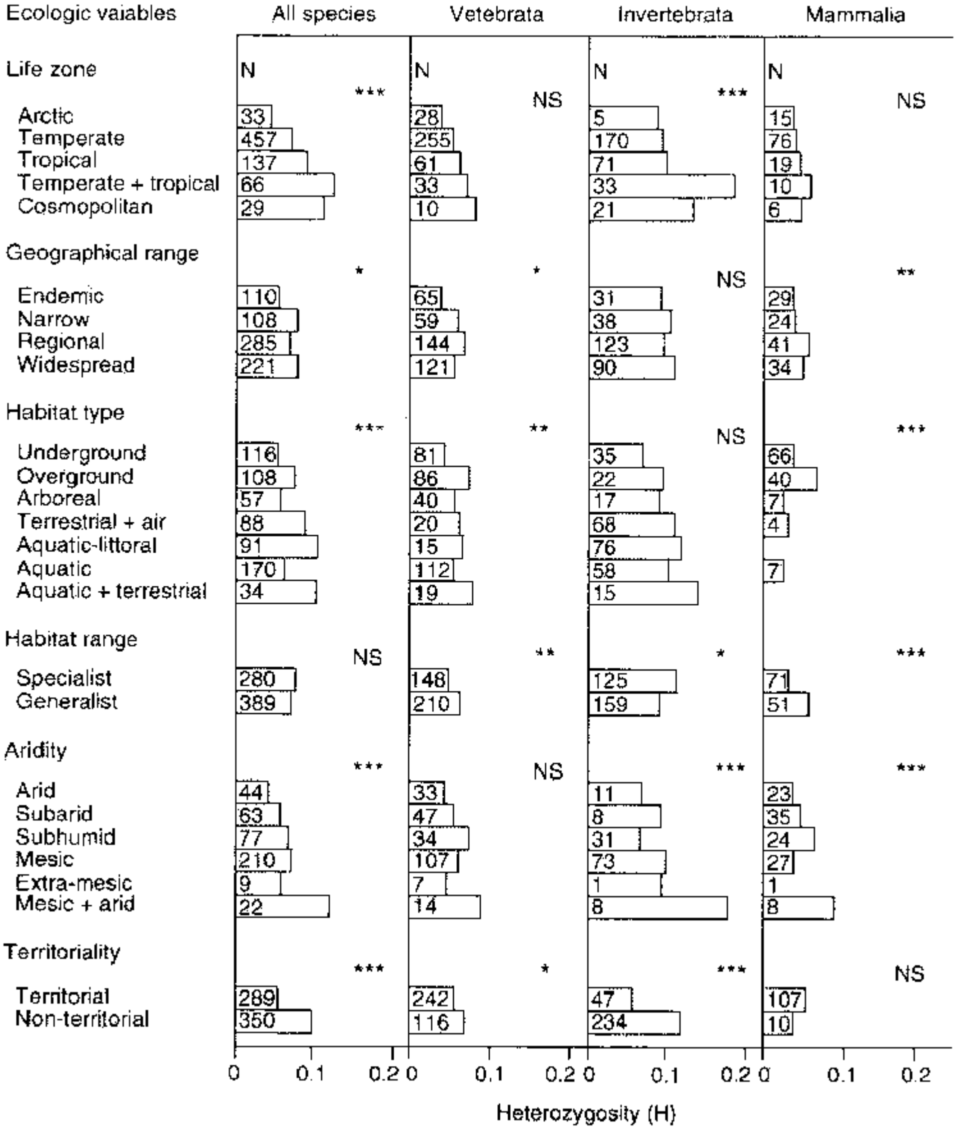


FIGURE 2 Global analysis of genetic diversity: levels of heterozygosity of ecological factors. Numbers in the bars indicate species. Significance ns = p > 0.05; * = p < 0.05; ** = p < 0.01; *** = p < 0.001 (reproduced with permission from Nevo *et al.*, 1984).

diverse animal species (Mitton, 1997). As expected, theoretically the global evidence exhibits a positive relationship between environmental diversity and genetic diversity, as was also shown in cage experiments. Clearly, environmental diversity enhances genetic diversity. Levins (1968) predicted that environmental grain determines the level of polymorphism. Fine-grained species (highly mobile) would evolve a monomorphic strategy, whereas coarse-grained species (highly sedentary) would maintain polymorphism. This theoretical prediction is largely supported by evidence (Nevo *et al.*, 1984).

The patterns and correlates of genetic diversity revealed in the global analysis for many unrelated species, subdivided into different abiotic and biotic regimes, strongly implicate selection in the genetic differentiation of species. Natural selection in several forms, but most likely through the mechanisms of spatiotemporally varying environments at the various life cycle stages of organisms, appears to be an important evolutionary force causing change at the molecular level. Other evolutionary forces, including mutation, migration, and genetic drift, certainly interact with natural selection, either directly or indirectly, and thereby contribute differentially, according to circumstances, to population genetic differentiation at the molecular level. However, the final orientation of the evolutionary process is determined by natural selection (Bell, 1997).

D. Regional Analysis of Genetic Polymorphisms

1. Regional: Allozyme Diversity across Phylogeny in Israel and the Near East

a. Israel as an Ecological Genetic Laboratory of Increasing Aridity Southwards

Our regional analyses of genetic diversity extended over Israel and the Near East. We used Israel, with its remarkable physical and biotic diversities, as a medium- to large-scale ecological genetic laboratory (Nevo, 1988, 1998). In 1988, Nevo and Beiles conducted an ecological test of protein polymorphism in 13 unrelated taxa of plants, invertebrates, and vertebrates involving 21 species, 142 populations, and 5474 individuals (Fig. 3), following the extensive studies of allozyme diversity in 38 species in Israel.

Each individual, population, and species were tested, on average, for 27 enzymatic gene loci. These species varied in population size and structure, life histories, and biogeographical origins, but they largely shared geographically short and ecologically stressful gradients of increasing aridity in Israel, both eastward (70 km) toward the Syrian and Jordanian deserts and (mainly) southward (260 km) towards the Negev desert. We found genetic parallelism across most taxa and most loci. Observed average heterozygosity, *H*, and gene di-

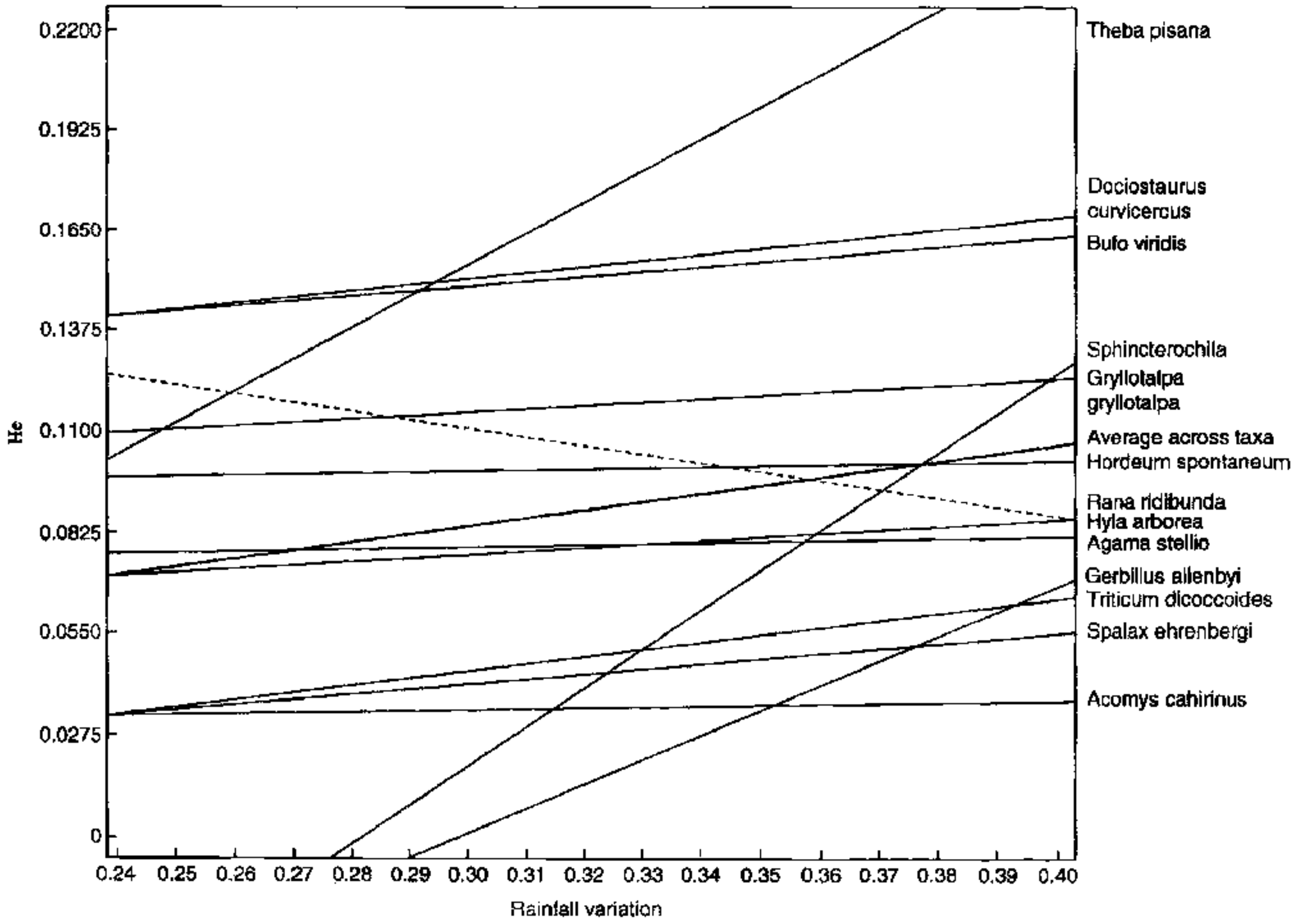


FIGURE 3 Parallel genetic patterns in the level of gene diversity, H_e, of 13 enzymatic systems averaged across all taxa studied here. The average regression line is $H_e = -0.0556 + 0.4803$ RV. The only change is that *Gryllotalpa gryllotalpa* has been divided into two species, *G. tali* ($2n = 19$) and *G. marismortui* ($2n = 23$). See Broza *et al.*, 1998, cited in Nevo, 1999. The regression line represents *G. tali* [reproduced with permission from Nevo and Beiles (1988)].

versity, H_e, were positively and overall significantly correlated with rainfall variation, which is lowest in the northwest and highest in southern deserts (i.e., with a wider, fluctuating, climatic temporal niche in the desert) (Fig. 3). Similar trends were found for several DNA systems in subterranean mole rats in Israel (Nevo, 1999). This result corroborates the environmental theory of genetic diversity, which regards ecological heterogeneity and stress as major determinants enhancing genetic divergence.

Notably, in the regional Israeli studies, heterozygosity, H, is negatively correlated with effective population size, N_e; H increases, whereas N_e declines drastically toward the deserts. This contradicts a basic postulate of neutrality theory, which predicts positive correlation between H and N_e (Kimura, 1983). Our results are clearly inconsistent with the genetic drift theory postulated in 1931 by Sewall Wright or the neutral theory of molecular evolution, even in its milder form of near neutrality (Ohta and Gillespie, 1996). Our results suggest that natural selection, through environmental range and stress, in space or time or both appears to be an important genetic differentiating evolutionary force at the protein and DNA levels (i.e., driving molecular evolution) (Nevo, 1998). Similar results were obtained in extensive ecological genetic studies in wild wheat in 1989 by Nevo and Beiles and in subterranean mole rats across the Near East and Asia Minor (Nevo, 1999). Genetic diversity either in the inbreeding wild cereals aboveground or in outbreeding mole rats underground is primarily determined by climatic selection. Similar conclusions—that is, that natural selection on allozyme polymorphism may be intense with selection coefficients in the range of 0.1–0.8—have also been reached for many model organisms (Mitton, 1997). Remarkably, the abundant genetic diversity found in natural populations of plants and animals provides immense genetic resources for crop improvement in agriculture, biotechnology, and gene therapy in medicine.

b. Genetic Parallelism and Ecological Stress in Subterranean Mammals

Parallel genetic patterns of increasing diversities (primarily southward and secondarily eastward) were shown in 1996 by Nevo and colleagues in Israeli subterranean mole rats for diploid chromosome numbers, nuclear allozyme, mtDNA, RAPDs, minisatellites, chiasma frequency, and organismal traits and also for approximately 200 genes by AFLP (in preparation). Clearly, genomic molecular evolution of nuclear and extranuclear diversities of proteins and DNA, coding and noncoding sequences, in subterranean mole rats, generally assumed to be neutral by proponents of the neutral theory of molecular evolution, is *nonrandom* and correlated positively with increasing aridity stress and climatic unpredictability. The increase in heterozygosity is primarily toward the southern and eastern deserts, where temporal stressful climatic fluctuations and unpredictability climax. Adaptive climatic selection appears to be the prime driving force in both molecular and organismal evolution of mole rats. Natural selection through fluctuating environmental stress (i.e., wider climatic niche) appears to be a major cause of molecular evolution and genome organization in subterranean mole rats, similar to the pattern found in aboveground organisms facing stressful unpredictable climates, across the same transect of increasing aridity stress (Nevo, 1999).

E. Local Population Genetics Studies across Sharp Ecological Contrasts

1. Past Microgeographic Studies in Plants and Animals

Microgeographical studies in nature, particularly those involving sharp contrasts (climatic, thermal, edaphic, geologic, topographic, and chemical), provide remarkable long-term experiments, representing small-scale natural laboratories, for analyzing ecologic genetic population dynamics. Since 1975, we have conducted several microgeographic studies in nature (Nevo, 1998). We employed thermal (high vs low temperature) and chemical (polluted vs clean areas) stresses in marine balanids and conducted an extensive research program on inorganic and organic pollution and its effect on allozyme diversity in marine organisms. Likewise, we examined edaphic (terra rossa vs basalt soil and rock vs deep soil) and microclimatic (sun vs shade and high vs low solar radiation) stresses in wild barley (*Hordeum spontaneum*), wild emmer wheat (*Triticum dicoccoides*), and *Aegilops peregrina*. The conclusions of these studies all point inferentially to natural selection as a major differentiating factor of qualitative and quantitative patterns of genetic diversity at single loci, but primarily at multilocus structures and genome organization (Nevo, 1998, 1999).

The following is the summary of our extensive studies in wild cereals (wild barley and wild emmer wheat) that have been studied spatiotemporally since 1975 at the Institute of Evolution at the protein and DNA levels on micro- and macrogeographical scales. We found significant climatic and soil divergence at four microsites as well as in our macrogeographic studies in Israel and the Near East (Nevo, 1998). All the foregoing studies illustrate massive genetic *nonrandom* divergence at sin-

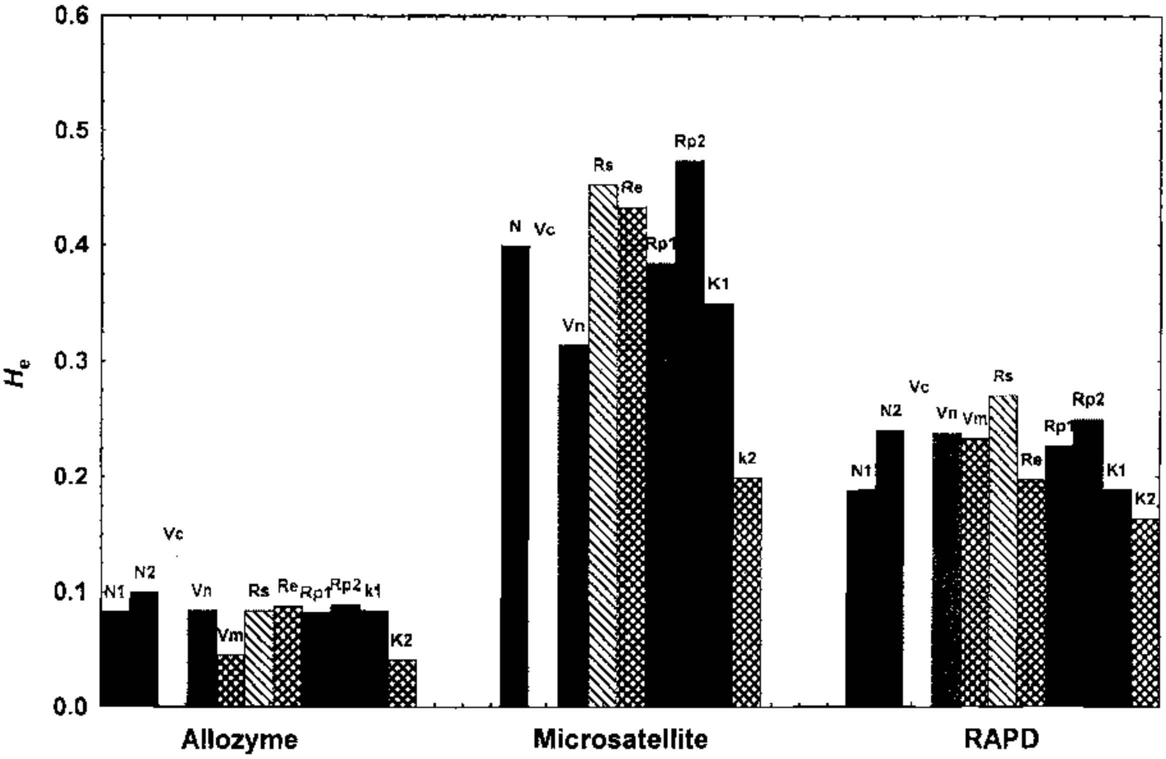


FIGURE 4 Genetic diversity, H_e, of wild emmer wheat, *Triticum dicocoides*, in four habitats (N, north; V, valley; R, ridge; K, Karst) divided into 11 microhabitats in the Ammiad microsite, Upper Galilee, as reflected in allozymes, microsatellites, and RAPDs (Li *et al.*, 2000a). Note the much higher levels of H_e in microsatellites and RAPDs compared with allozymes; likewise, note the similarity of all systems, but primarily note the high and significant correlation between microsatellites and RAPDs (r_s = 0.85, p = 0.003; see text). VC, main valley center; VM, main valley margins; VN, narrow valley; N_1, upper north-facing (moderate) slope; N_2, middle north-facing (steep) slope; RE, ridge, east-facing slope; RP_1, ridge, shoulder of plateau; RP_2, ridge, top of plateau; RS, ridge, south-facing slope; K_1, upper karst; K_2, lower karst (from Li *et al.*, 2000a).

gle-locus, two-locus, and multilocus structures of genome organization, with specific and unique alleles and levels of genetic diversity correlated with ecological heterogeneity, niche breadth, stress, and the largely parallel patterns of allozymes, RAPDs, and microsatellites despite increasing differences in levels of genetic diversity in this order (Fig. 4). Recently, we have shown the operation of edaphic selection on microsatellites across the microsites of Yehudiyya, Tabigha, and Ammiad (Fig. 5). Neither random drift nor gene flow can explain the levels and divergence of genetic diversity at the protein level or at the coding and noncoding DNA regions. The data strongly suggest that natural selection overrides the homogenizing effect of gene flow, maintains molecular polymorphisms in accordance with divergent ecologies, and orients molecular evolution. How general is this genetic divergence pattern of wild cereals, which are predominantly selfers, across life as a whole, primarily in outcrosser plants and nonsedentary animals?

In 1992, we embarked upon a long-term multidisciplinary project at the "Evolution Canyon" microsite. It aims to elucidate the causes of biodiversity and genome evolution at the molecular and organismal levels in diverse groups of organisms across phylogeny from cyanobacteria to mammals, representing a microcosm of life.

2. Evolution in Action across Phylogeny Caused by Climatic Stresses: "Evolution Canyon"

Biodiversity and genome evolution, and the relative importance of forces driving evolution, are critically tested at the "Evolution Canyon" microsite (Lower Nahal Oren, Mount Carmel), a dynamic ongoing microscale research program (Nevo, 1997). Our aim is to draw generalizations across life in genotyes and phenotypes and to highlight controversial and unresolved problems of biological evolution. The opposed slopes of "Evolution Canyon" display dramatic physical and

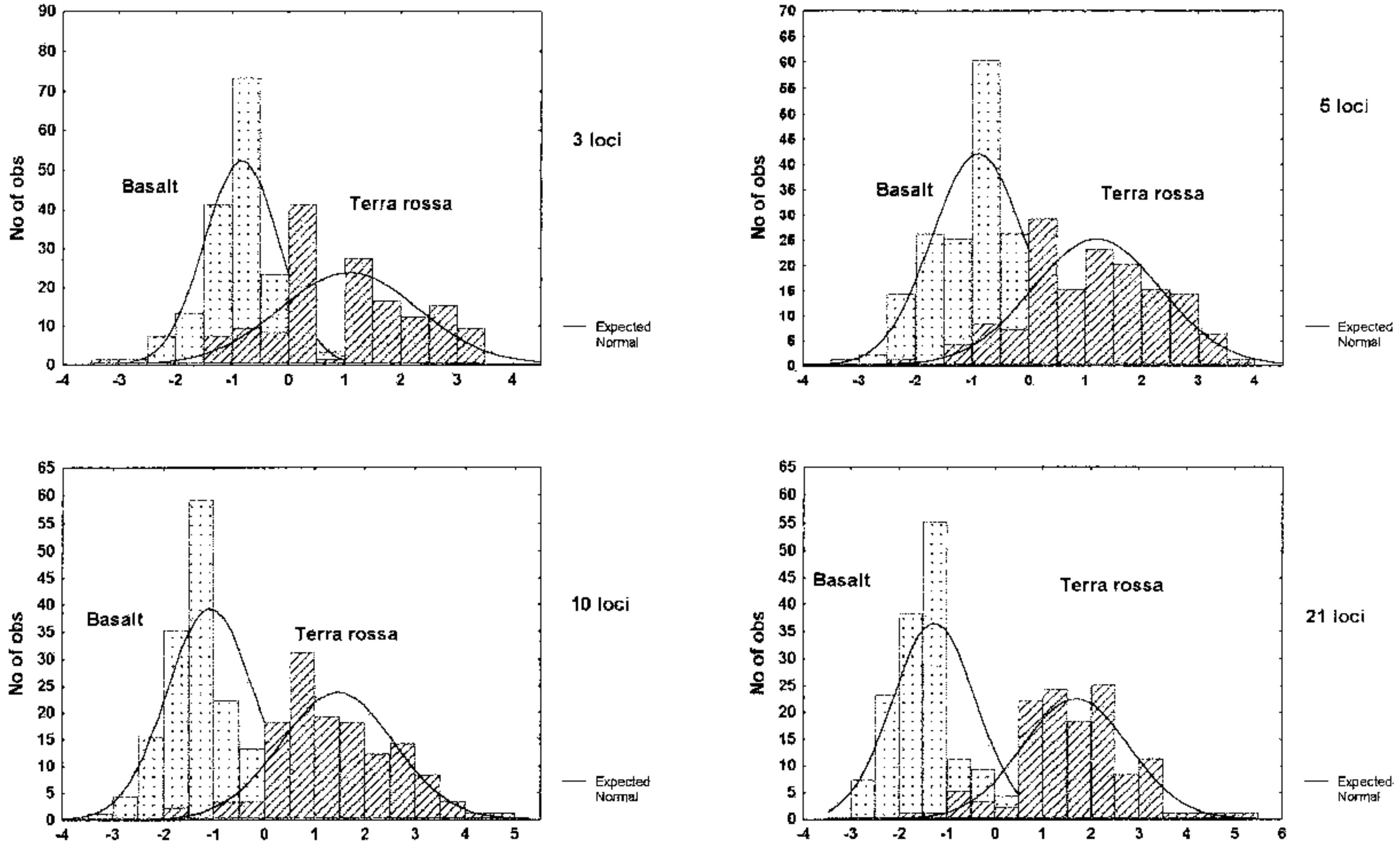


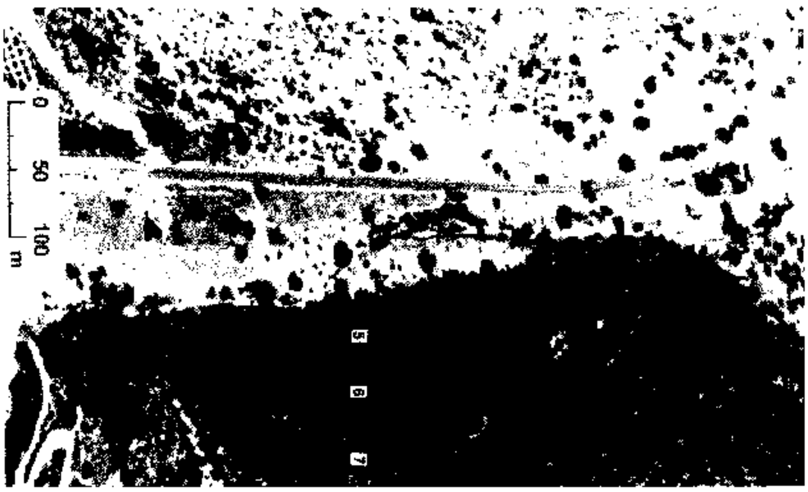
FIGURE 5 Discriminant analysis between two soil types, basalt and terra rossa of wild emmer wheat (*Triticum dicoccoides*), across three microsites (Yehudiyya, Tabigha, and Ammiad) north of the Lake of Galilee, Israel. This analysis is based on 2, 5, 10 and microsatellite loci, classifying correctly 21, 83, 88, 90, and 91% of plants into their original soil types (based on data from Li *et al.*, 2000b).

biotic contrasts on a microscale (Fig. 6). Higher solar radiation (up to 600% more) on the south-facing slope (SFS) makes it warmer, drier, and spatiotemporally more heterogeneous and fluctuating than the north-facing slope (NFS). The two slopes are separated by only 100 m (at bottom) and 400 m (at top). The SFS represents an Afroasian savanna park forest, whereas the NFS represents a dense Euroasian live oak maquis forest.

The tropical Afroasian warm and xeric SFS savanna displays wider ecological heterogeneity and higher stress for temperate terrestrial organisms. As predicted, genetic diversity was higher on the ecologically more heterogeneous (wider niche) and stressful SFS in 9 of 11 tested temperate model species (lichen, wild barley, 2 species of land snails, earthworm, diplopod, 3 beetles, and 2 rodents) (Fig. 7). In 6 species, heterozygosity was negatively correlated with population size, i.e., population abundance was lower, whereas genetic diversity was higher on the SFS. This negates the positive correlation expected between H and N_e by neutrality theory (Kimura, 1983). We have shown in wild barley, *H. spontaneum*, that RAPD and sequence-tagged site PCR analysis mirror allozyme interslope patterns.

Remarkably, heritable mutation and recombination rates were several-fold higher in the soil fungus *Sordaria fimicola* and in the fruit fly *D. melanogaster* on the more stressful SFS. Importantly, in terrestrial taxa species

Cross section



Air view

FIGURE 6 "Evolution Canyon," Lower Nahal Oren, Mount Carmel, Israel. Note the plant formation on opposed slopes. The lush, green "European" live oak maquis forest on the temperate, cool-mesic, north-facing slope sharply contrasts with the open park forest of warm-xeric, tropical "Afroasian" savanna on the south-facing slope (SFS). (Top) Cross section; (bottom) air view with the seven experimental stations—three on the SFS (1–3), one at the valley bottom (4), and three on the NFS (5–7) (reproduced with permission from Nevo, 1997).

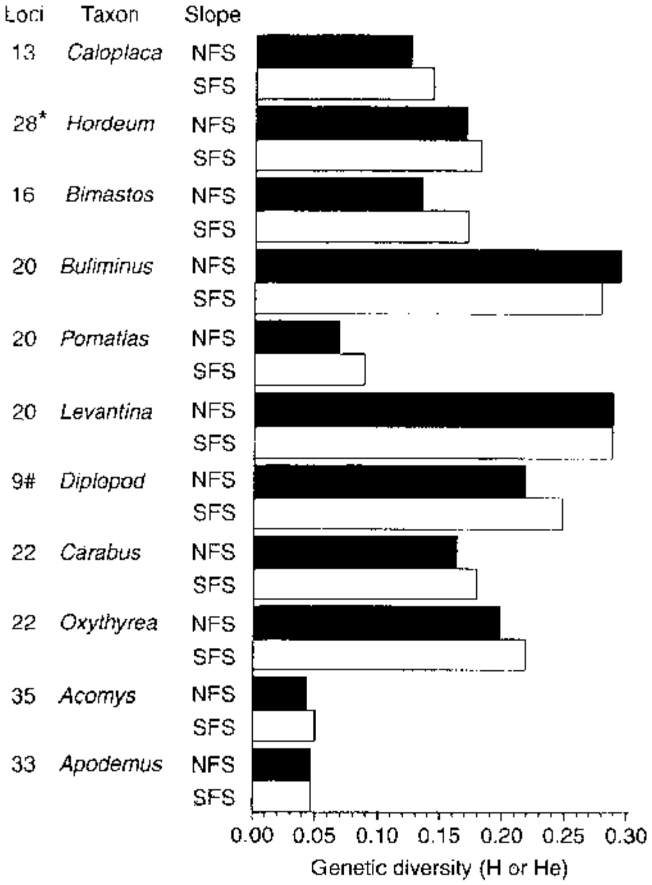


FIGURE 7 Genetic (allozyme) diversity in major plant and animal taxa on the opposed slopes of "Evolution Canyon"—the mesic, temperate, and mild north-facing slope (NFS) and the xeric, tropical, and stressful south-facing slope (SFS). *Caloplaca aurantia* (lichen), *Hordeum spontaneum* (wild barley), *Bimastos syriacus* (earthworm), *Buliminus labrosus*, *Pomatias olivieri*, and *Levantina caesareana* (land snails), *Tetrarthrosoma syriaca* (diplopod), *Carabus hemprichi* and *Oxythyrea noemi* (beetles), and *Acomys cahirinus* and *Apodemus mystacinus* (rodents) are presented. The genetic diversities represent allozyme heterozygosity observed (H) or expected (H_e). *Comparison between the highest stations (H_e); #, *Tetrarthrosoma syriacum* (reproduced with permission from Nevo, 1997)

richness was higher on the SFS, and interslope adaptive complexes were demonstrated in land snails, *Drosophila*, wild barley, and mice, contributing to differential fitness in morphology, physiology, behavior, and life history. Transplant experiments showed inter- and intraslope-specific fitness components matching the differential ecological stresses (all studies mentioned in this section are cited in Nevo, 1997). Similar results abound in local genetic differentiation in plants (Linhart and Grant, 1996).

We hypothesized that microclimatic diversifying natural selection, overriding migration and stochasticity, appears here as a major evolutionary driving force of genotypic and phenotypic evolution. The "Evolution Canyon" microsite provides a fertile microgeographic critical testing model of biodiversity evolution due to its sharp microscale ecological slope contrasts, both in niche width and in climatic stress.

F. The Nature of Allozyme and DNA Diversities

Critical laboratory experiments on marine model organisms have shown that allozymic diversity is adaptive and responds rapidly to environmental inorganic and organic pollution stresses (Nevo, 1998); a large amount of allozymic diversity can be gained or lost rapidly. Our experiments demonstrated unequivocally that ecological stresses of thermal and chemical pollution affect the level and pattern of allozyme allele frequencies and heterozygosity of marine organisms, thereby corroborating the environmental theory of genetic diversity. The survivorship values of heterozygotes and homozygote allozyme genotypes vary in accordance with the pollutant concentration. The broader the ecological niche, the higher the heterozygosity. Furthermore, heterozygote sites mutate more frequently than equivalent homozygous sites, possibly because mismatch repair between homologous chromosomes during meiosis provides extra opportunities to mutate. At medium levels of pollution, heterozygotes seem to be superior, whereas at high pollution levels specific homozygotes become superior. Our laboratory results on mercury pollution were confirmed in the sea. Remarkably, dynamic and rapid changes in genome size mediated by an increase or decrease in copy number of the retrotransposon BARE-1 in wild barley, *H. spontaneum*, both regionally and locally in Israel were described in 1999 by Schulman and colleagues.

G. Evolutionary Significance of Molecular Polymorphisms: Summary and Evidence

The foregoing macro- and microscale field and laboratory evidence across phylogeny displays massive genetic parallelisms to environmental heterogeneity and diverse stresses, physical (thermal, climatic, and edaphic) and biotic (pathogens, competitors, vegetation, and habitats). In general, higher levels of genetic polymorphisms, at both the protein and the DNA levels, occur under more stressful and variable environments. This is highlighted by the enormous diversity recorded in the vertebrate major histocompatibility complex (Bodmer and Bodmer as cited in Wasser, 1999) and immunoglobulin gene families, resistance genes in plants, against the immense pathogen diversity; and in the regulation of clock genes. Linkage disequilibria and genome organization (Nevo, 1998) are correlated with higher environmental heterogeneity (also called the niche width variation hypothesis by Van Valen, which predicts positive correlation between niche width and morphologic variation) and stress (Figs. 3 and 7). Similarly, mutation and recombination rates in the soil fungus *Sordaria fimicola* and in male *Drosophila melanogaster* at "Evolution Canyon" were both higher on the ecologically more heterogeneous and stressful SFS. Ecological heterogeneity and stress appear to be major determinants of the level of genetic diversity in nature, as was demonstrated in numerous organisms (Mitton, 1997). Physical and biotic ecological variables are better predictors of genetic diversity than demographic variables of population and species size. In fact, the levels of genetic diversity are often negatively correlated with effective population size, negating neutrality theory predictions. Generally, genetic polymorphism is positively correlated with niche width, or the level of ecological heterogeneity and stress, as emerged from our global, regional, and local analyses and from critical laboratory experiments. A brief theoretical discussion follows.

IV. THEORY

A. Mechanisms of Generating Genetic Diversity

It is beyond the scope of this article to describe the mutation mechanisms that generate diversity. Suffice it to say that a host of mechanisms provide a permanent input of diversity into natural populations, including point mutations, deletions, additions, recombination, gene conversion, and mismatch repair. Mutation rates also vary dramatically among organisms and genes and under diverse ecological stresses. Evidence from bacterial genetics (Casjens, 1998) suggests that organisms can resist stress through mutational changes or acquisition of pre-evolved functions via horizontal transfer. Activating mutagenic response and inhibiting antimutagenic activities (e.g., mismatch repair) could be important in recruiting genetic diversity as an adaptation to stress. Stress-induced increases in mutation rates enhance genetic polymorphism, primarily in genomic "hot spots."

B. Significance of Genetic Diversity

The functional value of genetic diversity within populations has been confirmed in a series of studies. Clearly, this rich pool of diversity provides the resource for continuous selection of adapted genotypes. Reduced genetic diversity may not only compromise the capacity of an impacted population for genetic adaptation in the face of further environmental challenge but also may result in increased energy requirements, lower production efficiency, and reduced homeostasis and reproductive output. These metabolic consequences of reduced genetic polymorphism would further lower that population's potential for survival under lethal conditions of contaminant exposure and also affect the genetic makeup of populations through differential reproduction under conditions of sublethal stress. In laboratory studies with population cages, higher levels of allozyme and additive genetic diversity are generally maintained in cages with greater heterogeneity, as is also generally true in nature (Mitton, 1997). Genetic diversity makes it possible to establish genetic distances within and between species, identify species and strains by unique and specific genetic profiles, establish phylogenetic trees, reinforce conservation biology and management programs, and have profound implications for plant and animal husbandry in agriculture and medicine.

C. Maintenance of Genetic Diversity in Nature

Explaining the maintenance of genetic diversity in natural populations has been a central problem of evolutionary genetics since the discovery of abundant protein polymorphisms in nature (Lewontin, 1974; Nevo, 1978, 1988, 1998; Kimura, 1983; Gillespie, 1991; Avise, 1994; Mitton, 1997). It is now clear that stochastic and bottleneck explanations, relating heterozygosity primarily to population size effects, are not realistic. Populations are rarely at equilibrium because environmental conditions fluctuate with abiotic (e.g., climate and seasons) and a host of biotic (e.g., parasites, pathogens, and competitors) factors. Correspondingly, genetic diversity can be gained or lost rapidly in dynamically changing natural populations due to ecological stresses. The major models explaining genetic diversity in nature relate primarily to the levels of ecological heterogeneity (niche breadth) and stress, or multiniches, and to environmental grain (Levins, 1968) rather than to effective population size or gene flow. Environmental variability enhances genetic diversity. Natural selection appears to be a major driving evolutionary force maintaining genetic diversity in nature (Bell, 1997). However, heterosis (i.e., overdominance or heterozygote advantage) alone is not a mechanism for maintaining many alleles segregating at a locus, as argued in 1978 by Lewontin, Ginzburg, and Tuliapurkar; hence the resort to spatiotemporal driving forces at multilocus structures.

D. Natural Selection

Theoretically, spatial and temporal variations of selection ("diversifying selection") could maintain and enhance genetic polymorphisms, although the conditions of their applicability are strongly limited [see cited papers by Levene, Haldane, Karlin, Lande, Felsenstein, Hedrick, Hoekstra, Maynard Smith in Ewens (1979) and Nevo (1998)]. Spatial variation appeared more effective than temporal variation, although their joint action could reinforce the maintenance of polymorphism. Most results related to selection of variation in time were derived from the one-locus case. Polymorphism maintenance may be reinforced in the case of two-locus or multilocus structures (Gillespie, 1991; Kirzhner *et al.*, 1998). Fitness components (viability, growth rate, fecundity, mating success, and developmental stability) increase with heterozygosity, suggesting that selection balances diversity at protein polymorphisms (Mitton, 1997). Some of the correlations between heterozygosity and fitness components can be attributed to allozyme loci or loci in strong linkage disequilibrium with them. This positive relationship between heterozygosity and fitness is not only expected theoretically but also abundantly documented empirically (Mitton, 1997). The selective mechanism is much more effective in promoting genetic diversity if carriers of the alternative alleles are able to select the niche in which their fitness is greatest as argued in 1977 by Taylor and Powell and by Nevo and colleagues in 2000 for mole crickets. For selection theories of molecular evolution and polymorphism based on the combined forces of natural selection, genetic drift, and mutation, see Gillespie (1991). For a discussion on the development of neutral and nearly neutral theories, see Ohta and Gillespie (1996). A series of articles dealing with natural selection and Darwinian fitness in nature appear in Wasser (1999).

In a series of articles and books during the period 1931–1978, Sewall Wright developed his "adaptive landscape" theory. In this model, populations are imagined to spend most of their time on selective peaks, with genetic drift providing the push when a population jumps from one peak to another. In contrast, Fisher's (1958) model on the nature of adaptation assumes that a population is never exactly at its optimal phenotype. Likewise, in Gillespie's (1991, pp. 291–305) random environment selection theories, there are no analogs to adaptive peaks because the adaptive landscape is changing faster than the genetic system: "The population is always running uphill, but the peak is always two steps ahead. All the population ever sees, in effect, is the side of the mountain. Should it stop evolving, it will face extinction." This view is related to Fisher's model and to the red queen hypothesis of Van Valen (1973), which assumes that if a population does not continue to adapt at the same rate as its competitors or environmental deterioration it may become extinct.

E. Stabilizing Selection in Cyclical Environments

Stabilizing selection for an intermediate optimum is generally considered to rapidly deplete genetic diversity in quantitative traits, with increased number of loci. In contrast to previous conclusions, we found in both haploids and diploids, in the case of an additive two-locus model, that stabilizing selection with cyclically moving optimum may be an efficient factor in protecting polymorphisms for linked loci additively affecting the selected trait (Korol *et al.*, 1994; Kirzhner *et al.*, 1998). We proved that within the same class of fitness functions, nonequal gene action and/or dominance effect for one or both loci may lead to local polymorphism stability with substantial polymorphism-attracting domain. A higher intensity of selection could result in two forms of polymorphic limiting behavior: (i) the usually expected forced cycles with a period equal to that of environmental changes and (ii) "supercycles," which are nondamping autooscillations with a period composed of hundreds of forced oscillations.

We have demonstrated (Kirzhner *et al.*, 1998) that a multilocus system subjected to stabilizing selection with cyclically moving optimum can generate ubiquitous complex limiting behavior, including supercycles, T cycles, and chaotic-like phenomena. This mode of multilocus dynamics far exceeds the potential of complex dynamics attainable under ordinary selection models resulting in simple behavior. It may represent a novel evolutionary mechanism increasing genetic polymorphism over long-term periods (Kirzhner *et al.*, 1998).

F. Selection versus Genetic Drift in Small Isolated Populations

Remarkably, our findings of high-level heterozygosity in small (several dozen or 100 individuals) desert isolates of the subterranean mole rats, *Spalax ehrenbergi* superspecies ($2n = 60$), in the northern Negev (Nevo,

1999) contradict both the Wrightian notion of genetic drift in small populations that was elaborated in 1931 and the extreme version of genetic drift, i.e., the neutral theory of molecular evolution (Kimura, 1983). The latter predicts positive correlation between effective population size (N_e) and heterozygosity *H*, whereas our finding demonstrated the opposite (i.e., the highest level of *H* in the smallest populations, or a negative correlation between *H* and N_e). A negative correlation between population size and heterozygosity was also found in several species tested at the "Evolution Canyon" microsite (Nevo, 1997). Current theoretical models predict fast gene fixation in small panmictic populations without selection, mutation, or gene inflow. Using simple multilocus models, we demonstrated that moderate stabilizing selection (with stable or fluctuating optimum) for traits controlled by additive genes could oppose random fixation in such isolates during thousands of generations (Nevo *et al.* 1997a, as cited in Nevo, 1999). We also showed that in selection-free models polymorphism persists only for a few hundred generations even under high mutation rates. Our multichromosome models challenge the hitchhiking hypothesis of polymorphism maintenance for many neutral loci due to close linkage with a few selected loci.

G. Genetic Ecological Diversity and Stress

The foregoing discussion suggests several explanations for the maintenance of genetic diversity subjected to ecological diversity and environmental stress. Spatial and temporal ecological variation, which predominates in nature, is of prime importance in maintaining and enhancing genetic diversity in natural populations. This may be true because different genotypes display varying fitnesses in variable environments and stresses. Recombination frequencies and mutation rates tend to increase under stressful conditions (Hoffmann and Parsons, 1991; Korol *et al.*, 1994; Nevo, 1997). Rates of evolutionary change are therefore enhanced in adverse environments, as was demonstrated under controlled laboratory experiments in the case of mercury pollution (Baker *et al.* as cited in Nevo, 1998), under regional aridity stress across the physically stressful Israeli environment, and locally at "Evolution Canyon" because of high solar radiation on the SFS (Nevo, 1997).

Developmental instability may be enhanced under both environmental stress and genomic stress. This could increase genetic diversity under stressful conditions (Hoffmann and Parsons, 1991). Heterozygote advantage tends to increase diversity with stress up to extreme points. However, heterosis alone is not a mechanism for maintaining many alleles segregating at a locus. It is much more likely that stable equilibria for multiple alleles will be best explained by multiple-niche selection. Models of sexual reproduction as an adaptation to resist parasites proposed by Hamilton and colleagues in 1990 may also contribute to sex evolution, recombination, and polymorphism. Finally, the simple model advanced by Kirzhner and colleagues in 1999 of genetic interaction between multiple species on a trait for trait basis governed by abiotic and biotic selection for multilocus quantitative traits opens wide horizons for the evolution of genetic diversity due to species dynamic interactions in nature.

Ecological heterogeneity and stress appear to enhance genetic polymorphisms, particularly in dynamically cycling environments (which can generate complex dynamic-like supercycles and chaotic-like behavior). This mode of multilocus dynamics far exceeds the potential for maintaining genetic polymorphism attainable in ordinary selection models. It may represent a novel evolutionary mechanism increasing genetic polymorphism over long-term periods. This novel mechanism could contribute to the observation that biological diversity has increased over geological time despite the well-known massive extinctions, providing ever-increasing genetic diversity and thus enhancing the evolution of biodiversity.

Acknowledgments

I thank Abraham Korol and Avigdor Beiles for critically reading and commenting on the manuscript. This study was supported by the Israel Discount Bank Chair of Evolutionary Biology and the Ancell–Teicher Research Foundation for Genetics and Molecular Evolution.

See Also the Following Articles

ECOLOGICAL GENETICS • EVOLUTION, THEORY OF • GENES, DESCRIPTION OF • NUCLEIC ACID BIODIVERSITY • PHENOTYPE, A HISTORICAL PERSPECTIVE

Bibliography

Avise, J. C. (1994). *Molecular Markers, Natural History and Evolution*. Chapman & Hall, New York.

Bell, G. (1997). *Selection. The Mechanism of Evolution*. Chapman & Hall, New York.

Bodmer, W., and McKie, R. (1994). *The Book of Man. The Quest to Discover Our Genetic Heritage*. Abacus, London.

Casjens, S. (1998). The diverse and dynamic structure of bacterial genomes. *Annu. Rev. Genet.* **32**, 339–377.

Curole, A. P., and Kocher, T. D. (1999). Mitogenomics: digging deeper with complete mitochondrial genomes. *Trends in Ecology and Evolution* 14, 394–398.

Darwin. C. (1859). *On the Origin of Species by Means of Natural Selection.* Murray, London.

Ewens. W. J. (1979). *Mathematical Population Genetics.* Springer-Verlag, Berlin/New York.

Fisher, R. A. (1958). *The Genetical Theory of Natural Selection.* Dover, New York.

Gillespie, J. H. L. (1991). *The Causes of Molecular Evolution.* Oxford Univ. Press, Oxford.

Goldstein, D. B., and Schlotterer, C. (1999). *Microsatellites: Evolution and Applications.* Oxford Univ. Press, London.

Hamrick, J. L., Linhart, Y. B., and Mitton, J. B. (1979). Relationships between life history characteristics and electrophoretically detectable genetic variation in plants. *Annu. Rev. Ecol. Syst.* 10, 173–200.

Hoffmann, A. A., and Parsons, P. A. (1991). *Evolutionary Genetics and Environmental Stress.* Oxford Univ. Press, Oxford.

Johns, G. C., and Avise, J. C. (1998). A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Molecular Biology and Evolution* 15, 1481–1490.

Karlin, S., Campbell, A. M., and Mrazek, J. (1998). Comparative DNA analysis across diverse genomes. *Annu. Rev. Genet.* 32, 185–225.

Kimura, M. (1983). *The Neutral Theory of Molecular Evolution.* Cambridge Univ. Press, Cambridge, UK.

Kirzhner, V. M., Korol, A. B., and Nevo, E. (1998). Complex limiting behaviour of multilocus genetic systems in cyclical environments. *J. Theor. Biol.* 190, 215–225.

Korol, A. B., Preygal, I. A., and Preygal, S. L. (1994). *Recombination Variability in Evolution—Algorithms of Estimation and Population Genetics Models.* Chapman & Hall, London.

Levins, R. (1968). *Evolution in Changing Environments.* Princeton Univ. Press, Princeton, NJ.

Lewontin, R. C. (1974). *The Genetic Basis of Evolutionary Change.* Columbia Univ. Press, New York.

Li, Y. C., Krugman, T., Fahima, T., Beiles, A., Röder, M. S., Korol, A. B., and Nevo, E. (2000a). Parallel microgeographic patterns of genetic diversity and divergences revealed by allozyme, RAPD, and microsatellites in *Triticum dicoccoides* at Ammiad, Israel. (submitted).

Li, Y. C., Fahima, T., Korol, A. B., Peng, J. H., Röder, M. S., Kirzhner, V., Beiles, A., and Nevo, E. (2000b). Microsatellite diversity correlated with ecological-edaphic and genetic factors in three microsites of wild emmer wheat in north Israel. *Mol. Biol. Evol.* 15(5).

Linhart, Y. B., and Grant, M. C. (1996). Evolutionary significance of local genetic differentiation in plants. *Annu. Rev. Ecol. Syst.* 27, 237–277.

Marra, M. *et al.* (1999). An encyclopedia of mouse genes. *Nature Genetics* 21, 191–194.

Martin, W., Stoebe, S., Goremykin, V., Hansmann, S., Hasegawa, M., and Kowallik, K. (1998). Gene transfer to the nucleus and the evolution of chloroplasts. *Nature* 393, 162–165.

Mitton, J. B. (1997). *Selection in Natural Populations.* Oxford Univ. Press, Oxford.

Nevo, E. (1978). Genetic variation in natural populations: Patterns and theory. *Theor. Population Biol.* 13, 121–177.

Nevo, E. (1988). Genetic diversity in nature: Patterns and theory. *Evol. Biol.* 23, 217–247.

Nevo, E. (1997). Evolution in action across phylogeny caused by microclimatic stresses at "Evolution Canyon." *Theor. Population Biol.* 52, 231–243.

Nevo, E. (1998). Molecular evolution and ecological stress at global, regional and local scales: The Israeli perspective. *J. Exp. Zool.* 282, 95–119.

Nevo, E. (1999). *Mosaic Evolution of Subterranean Mammals: Regression, Progression and Global Convergence.* Oxford Univ. Press, Oxford.

Nevo, E., Beiles, A., and Ben-Shlomo, R. (1984). The evolutionary significance of genetic diversity: Ecological demographic and life history correlates. *Lect. Notes Biomath.* 53, 13–213.

Ohta, T., and Gillespie, J. H. (1996). Development of neutral and nearly neutral theories. *Theor. Population Biol.* 49, 128–142.

Ramshaw, J. A. M., Coyne, J. A., and Lewontin, R. C. (1979). The sensitivity of gel electrophoresis as a detector of genetic variation. *Genetics* 93, 1019–1037.

Schena, Heller, A. R., Theriault, T. P., Konrad, K., Lachenmeier, E., and Davis, R. W. (1998). Microarrays: biotechnology's discovery platform for functional genomics. *Trends in Biotechnology* 16, 301–306.

Van Valen, L. (1973). A new evolutionary law. *Evol. Theory* 1, 1–30.

Wasser, S. P. (ed.) (1999). *Evolutionary Theory and Processes: Modern Perspectives.* Kluwer, Dordrecht.

Watson, J. D., Hopkins, N. H., Roberts, J. W., Steitz, J. A., and Weiner, A. M. (1987). *Molecular Biology of the Gene*, Vols. 1 and 2. Benjamin/Cummings, Menlo Park, CA.

GEOLOGIC TIME, HISTORY OF BIODIVERSITY IN

James W. Valentine
University of California, Berkeley

GLOSSARY

alpha diversity Diversity of species within a single habitat.

beta diversity Measure of the rate and extent of change in species along a given habitat or physiographic gradient.

disparity Range of diverse morphological architectures present in higher taxa such as classes, phyla, and kingdoms; it is used to explain the origins and maintenance of this diversity of life-forms and body plans.

taxonomic richness Number of species, genera, and families from a given time period or fossil excavation.

LEARNING THE PATTERNS OF PAST BIODIVERSITY CHANGE is fundamental to understanding the causes and consequences of processes that generate and deplete the diversity of life on earth. Although the fossil record is incomplete and biased, methods have been developed to correct for these problems in large measure, thus permitting the study of biodiversities of the geologic past.

I. BIODIVERSITY IN THE FOSSIL RECORD

There are two distinctive though related aspects to paleobiodiversity studies. One approach focuses on the number of taxa present during a certain geologic time, or over a certain geologic time span. The taxonomic categories involved are usually species, genera, or families, and this sort of diversity is commonly termed taxonomic *richness*. The ultimate aim of most such studies is to understand and explain the origin and maintenance of patterns of species richness in space and time; when genera or families are used it is usually as proxies for species. The evolution of species richness has been studied since Darwin's time and continues to be a highly productive area of research. The second aspect deals with diversity, not of numbers of species, but of different types of organisms. The taxonomic categories involved are the higher ones, such as kingdoms, phyla, and classes, and the aim is to understand and explain the origins and maintenance of *disparity*—the range of diverse morphological architectures that is present. The major patterns of disparity over geologic time have been described, chiefly using taxonomic categories, as proxies for morphological distance. However, the evolution of that disparity has received much less attention than the evolution of richness, partly because it was thought that disparity could be explained by extrapolation from the processes that create richness. With the rise of molecular developmental studies, however, the genetic basis for disparity is becoming understood, and it has become possible to decouple studies of the origin of disparity from those of richness.

The ecological regulation of local biodiversity is not well understood, but empirical patterns of species distributions and richnesses indicate some of the factors governing species accommodation on regional and global scales. These large scales are of particular importance to tracing biodiversity through time. A basic measure of species richness is *alpha diversity*, the number of species found at a given locality or in a single sample. A second sort of species richness, *beta diversity*, is a measure of the additional species that are found at a second locality or in a second sample. Alpha diversity was taken to indicate the amount of species packing (reflecting ecological niche partitioning) at a locality, and beta diversity to indicate the extent of habitat differences between localities. In practice, however, different species are found at a given locality at different times, varying with season, weather, or other factors that affect successful reproduction or colonization at the locality. Alpha diversity can therefore vary significantly from time to time, in which case beta diversity is affected as well.

The global environment is quite heterogeneous, a mosaic of habitats separated by gentle gradients in some places and by sharp boundaries in others and further broken into disjunct regions, such as different continental masses or remote islands. Beta diversity will vary according to whether two samples are from adjoining areas within the same general habitat type (i.e., represent the same biotic community), or are from distinctive habitat types (represent different communities), or are from distinctive climatic or physically disjunct regions (represent different biotic provinces). Though beta diversity is sensitive to differences among samples within a community, it is not as useful a measure between communities; rather, a more appropriate measure is the number of species added as entire communities are combined—a sort of super beta diversity. To capture global biodiversity patterns, a hierarchy of beta-type measures is required; differences between bioprovinces can be evaluated by the numbers of species added when combining one entire bioprovince with the next, or similar evaluations can be made between entire realms. The pattern of environmental heterogeneity in the world has varied continuously over geologic time, thus affecting the pattern of accommodation of biodiversity.

II. EVOLUTION AND BIODIVERSITY

A. Richness

Speciation is produced by isolation of populations, severing the flow of genes between them and permitting the evolution of differences in their features, underlain by differences in genes, gene frequencies, and gene organizations that arise between the populations. Once sister population crosses are unable to produce fertile offspring, speciation is complete. Environments favoring more isolation among populations are therefore more likely to produce rich biotas, other things being equal. A world broken into many continents favors global species richness, with a nearly separate biota on each continent. A continental landscape consisting of a great variety of climates and habitat types favors more species richness than does a monotonous landscape, for species tend to become well adapted to particular habitats, and the more habitat types available, the more species. A landscape in which any given habitat type is scattered among other types, which therefore provide barriers to gene flow, favors more species richness than a landscape where each habitat type occurs in a large tract. In short, environmental heterogeneity clearly promotes isolation and therefore species richness through speciation. Other factors also promote speciation, but they are not so clear nor so well understood. For example, low-latitude, tropical regions hold many more species than midlatitude temperate regions, and these latter hold more than high-latitude arctic/antarctic regions, but the mechanisms that link the obvious climatic differences to this latitudinal gradient in species richness are not yet understood. It is clear that the reduction of environmental heterogeneity by human activities is lowering the capacity of the globe to support species.

B. Disparity

The genetic differences producing morphological disparity are not based so much on differences in genes and gene frequencies as on gene organization. Individuals belonging to different higher taxa, such as phyla, differ from their very early stages, producing different adult body plans as they develop. In animals, the developmental regulatory genes that are responsible for organizing the many disparate body plans are surprisingly similar. The animal phyla originated over half a billion years ago, so similarities in those developmental control genes have been conserved since that time. Even members of other kingdoms, such as fungi and plants, share some similarities with animals in their developmental control genes, which therefore have been conserved for over a billion years. Some key developmental genes have been found in all phyla that have been investigated, though they may play a variety of roles. These genes are responsible for regulating still other genes. The regulatory genes govern cascades of gene expression that mediate cellular differentiation and organogenesis, the hallmarks of multicellular body plans, and control the final body architecture. The evolution of disparity has involved the evolution of these basic developmental control systems. Thus, although the genes are similar among phyla, many of the gene expression pathways are different within each phylum and class and so produce distinctive body plans.

In animals, the use of genes that mediate the development of numbers of cell types and morphologies seems to have originated in a spongelike organism. Animal genomes contain two to three times as many genes as most unicellular protists, suggesting that one or more early, massive gene duplication event(s) provided the genetic resources that were necessary to regulate the development of complex morphologies. Vertebrates have approximately four times as many genes as an average invertebrate, suggesting that some massive gene duplication events in early chordate lineages provided a genome of the requisite size to permit the evolution of the very complex vertebrate bodies. Regulatory genetic systems have evolved in land plants as well, quite independently, mediating the disparity found in plant anatomy, but as yet these systems are not as well known as those of animals.

III. CAPTURING FOSSIL BIODIVERSITIES

A. Problems of Sampling Fossil Biodiversity

It is not easy to become a fossil; the remains of only an exceedingly small fraction of all the organisms that have lived have survived to provide information on the nature of life through geologic time. Therefore we must treat the fossil record of biodiversity as a huge sampling problem. One approach would be to estimate how many samples from how many places it would take to adequately represent present biodiversity, and then to compare this ideal sampling pattern with the sorts of fossil samples with which nature has presented us. To estimate today's biodiversity we would need to sample each community in each province on each continent in each major realm, such as terrestrial plains and moutains, continental shelves, and each major oceanic water mass and deep-sea region. We would not have to find every last species, but we would have to understand the patterns of species' distributions well enough to calculate biodiversity from our samples. Although valiant attempts have been made, the problem is so large that estimates of modern biodiversity are not closely constrained. Assessing fossil samples is not easy either. We need to know how densely the biota has been sampled regionally and ecologically by nature in presenting us with the fossil assemblages. And for a given fossil assemblage, we must estimate how completely the living biota was sampled, without knowing how diverse it was in the first place. We also need to know how different ancient biotas were from region to region, for it is not uncommon for large regions to lack fossils for given time periods, and their diversities must then be estimated from assumptions about global patterns. Finally, we must be able to estimate the temporal density of the fossil record and make allowances for intervals of time that lack records.

B. Completeness of the Rock Record

The rocks from which fossils have been recovered are nearly all exposed on land and have therefore been subjected to erosion, and many fossiliferous formations have been stripped away and lost. The first question to ask when reconstructing fossil biodiversities is how much of the original rock record remains. The completeness of a sequence of rocks is independent of its age, relating rather to its depositional history. In general, sedimentary sequences contain many small gaps but increasingly fewer gaps of longer durations. For a given sedimentary sequence, then, fossil taxa with very short geologic ranges may have been present only during a gap, whereas longer-ranging forms can "jump" most gaps and may be captured as fossils. Fortunately, the gaps in one region are commonly represented by sediments elsewhere, though a few periods of time are poorly represented in most regions. For example, few marine sediments are preserved on continents from times when sea levels were particularly low, such as at the end of the Permian and beginning of the Triassic periods.

Terrestrial sedimentary rocks, deposited as they are above sea level, are subjected to very active processes of erosion, which remove much of the sediment in rivers or in winds and deliver it to the sea. Preservation of terrestrial formations is thus restricted chiefly to downwarped epicontinental basins or interior continental platforms, where there is some protection from the main erosional forces. Such settings tend to be more localized and geologically shorter-lived than nearshore marine sites. Furthermore, terrestrial animal remains are also exposed to decay and erosion in the subaerial environment and tend to be destroyed more easily than the shells of marine forms. Thus, terrestrial rock sequences are less complete than marine ones, and terrestrial animals are not as well sampled by the fossil record as marine animals.

C. Incompleteness of the Fossil Record

Organisms differ greatly in their potential for preservation as fossils. In general, entirely soft-bodied, or small, or rare, or geographically restricted species of short durations have fewer opportunities for preservation than their opposites: a large, abundant, widespread species that lived for many millions of years and has a very durable skeleton is most likely to be represented by fossils. Some phyla, such as Mollusca (clams and snails, etc.) are well represented in marine sediments, whereas others, such as Platyhelminthes (flatworms), though at least as old as the Mollusca and quite rich in species, have left no recognized body fossils. The biodiversity of the living biotas from which fossil assemblages are recruited must usually have been considerably greater than the numbers of fossil taxa that are found. However, it seems that most species with durable skeletons are in fact originally captured in sediments. For example, a study of fossilization of living species of marine bivalves (Mollusca) of the Californian bioprovince found that about 80% of the species, 84% of the genera, and 90% of the families were represented by fossils in sediments of Pleistocene age (the last 1.6 million years). As expected, the missing species were small, thin-shelled, and rare forms, and the missing genera and families were made up of such species.

In richness studies of taxa with high preservation potentials, such as shelled mollusks, fossil assemblages actually have an advantage over samples of the living biota. The fossils are usually time-averaged, that is, they represent the accumulation at a given locality of many generations and may include individuals of species that have occupied the region for only some fraction of the time represented. Thus, a fossil assemblage can sample the species richness that was found in a given area far better than a sample of alpha diversity in a living community. The fossils represent a sampling through time, often over hundreds to thousands of years.

Foote and Sepkoski (1999) devised a method that estimates the probability of preservation of fossil higher taxa from a sort of frequency distribution of the observed geologic range lengths of their members. When such an estimate is made for marine bivalves (clams), the probability of preservation of a genus within a stratigraphic interval representing about 5 million years is around 50% (Fig. 1). The bivalve faunas preserved in the sediments have been significantly degraded. Although some elements of the fossil faunas may have been destroyed by dissolution within the sediments, it seems likely that most of the missing genera were originally present in sediments that were then either removed during intervals of erosion or are not now exposed for collection and study. In other words, although fossils commonly represent a good sample of living associations of easily preservable taxa, not enough such associations are preserved to directly count ancient diversities over lengthy intervals of geologic time.

The completeness of the record of a number of other higher taxa is given in Fig. 1, where the probability of their generic preservation is plotted against the proportion of the living families that have a fossil record. There are two groups that deviate from the general trend. The Cephalopoda (Mollusca) fall well above the trend because they are represented by rich fossil assemblages of extinct higher taxa and have few living representatives. By contrast, Chondrichthyes (Chordata) fall well below the trend because their living families have a good fossil record, chiefly of teeth found in relatively young sediments, but the earlier fossil genera, represented by bodies, are mostly known only from single localities and thus have low preservation potentials. Thus the two outliers are readily explained, and the correlation otherwise evident between these totally independent data sets suggests that they are reasonably accurate. Because the marine sedimentary rocks of continental shelves combine high completeness and availability, the marine invertebrates with the highest preservation potentials provide the longest, most complete fossil record of any major groups of organisms.

In establishing the geologic range of a fossil taxon, it is almost certain that we have not found the earliest members to evolve or the last individuals to become

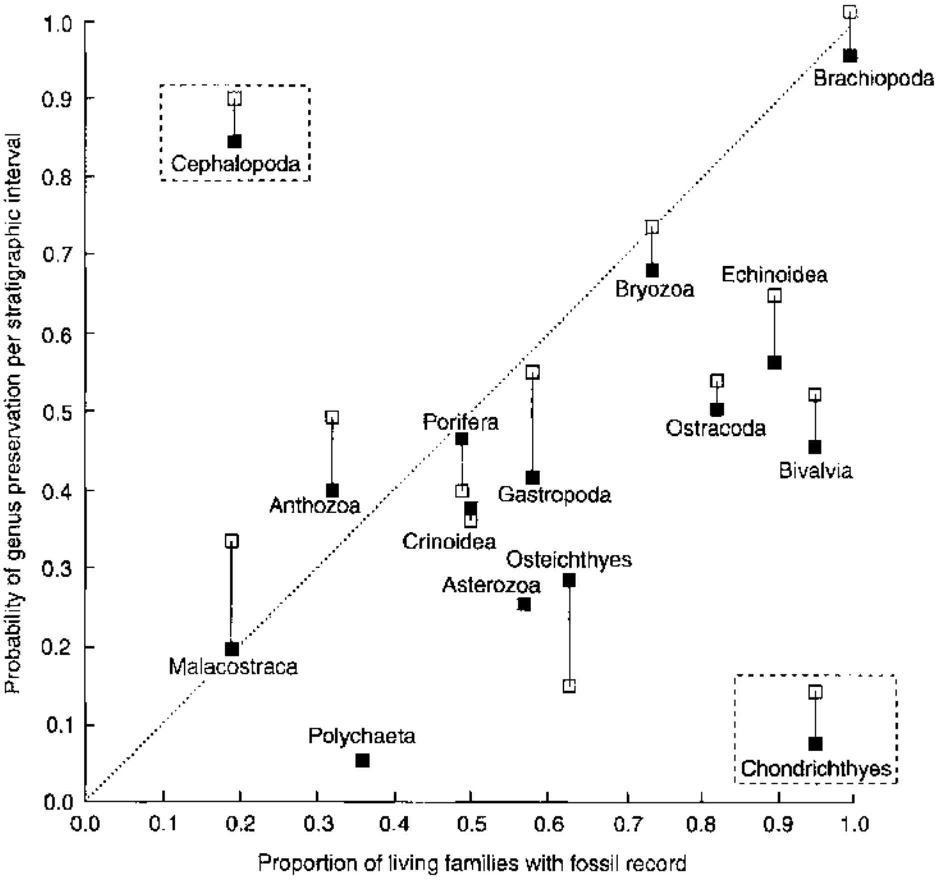


FIGURE 1 Two measures of completeness for taxa of marine animals: the vertical scale is the calculated probability of the preservation of a genus of the taxa per stratigraphic interval; the horizontal scale is the proportion of living families of the taxa with a fossil record (a measure of the preservation potential of the taxa). There are two points (boxes) for each taxon because two different timescales (with different bins) were used to evaluate binning biases. The significant correlation of these two measures indicates that the estimated completenesses are generally accurate. The Cephalopoda and Chondrichthyes are outliers; see text for discussion. (Reprinted with permission from *Nature* Foote and Sepkoski, © 1999 Macmillan Magazines Limited.)

extinct, because the fossil record is so incomplete. It is possible, however, to judge the probable extent of a taxon's range beyond its actual record by how commonly it is found between its first and last occurrences: if it occurs rather continuously within its known range, then when it disappears it is likely to be absent; if it occurs rarely, then it may have existed well before we see it and continued to exist well after our last record. The number of a taxon's occurrences in a given time interval thus permits a statistical estimate of the likely extent of its true geologic range. By taking account the observed preservation potentials of a taxon, estimates of the numbers of that taxon present in a given interval can be augmented by granting range extensions beyond the fossil occurrences. Corrections of this sort will produce better approximations of the richness within geologic intervals than the raw numbers of fossil taxa in collections.

D. Partial Remedies for the Spotty Fossil Biodiversity Record

The taxonomic hierarchy, based on Linnaean principles, produces many paraphyletic taxa, but it is useful for purposes of biodiversity reconstruction. Even within taxa with the highest preservation potentials, a large proportion have been lost to erosion and other factors. Genera, on the other hand, contain on the average several species, so that a genus is usually more widespread, occurs in more environments, and contains many more individuals than any one of its species. As a consequence, the fossil representation of genera is more complete than that of species. Families usually include several genera and so are even more completely represented as fossils. Thus by compiling biodiversity data on the generic level, one produces a significantly more complete and somewhat more accurate record than is possible at the species level, and the record at the family level is even better.

There is a price for this taxonomic remedy. The lower levels of taxonomic biodiversity are the more volatile through time, being more sensitive to the environmental changes that affect biodiversity. For example, a wave of extinction may decimate a biota at the species level, but to be recorded in a compilation of family diversity an extinction must extirpate every last species of a family. A small extinction wave might not even be noticed at the family level, considering that fossil data are rather noisy anyway. Of course the time of an extinction, whether recorded at the family, generic, or species level, will be registered in precisely the same place on the geologic timescale. Because species contain reproducing populations, they are in direct ecological contact with environmental parameters, and the volatility of their richness is the most sensitive measure of environmental changes. Higher taxa record the outcome of environmental change more indirectly, that is, when their species first appear, or when all of them disappear from the record. Nevertheless, their behavior in the face of environmental change presumably reflects the broad similarities in physiological and behavioral responses of their species, inherited from common ancestral species. While generic and family biodiversities can make quite useful proxies for species diversity in the fossil record, they are in fact measuring different features, and this should be taken into account in interpretations of fossil richness.

Most individual fossil samples represent a particular ecological community, perhaps somewhat intermixed with rare species that are chiefly found in other communities, where they are common. In some cases, though, the samples represent mixtures of common faunal elements that represent different communities but that lived in close proximity. It is sometimes possible to establish biodiversities for metacommunities by taking many samples of similar, approximately contemporaneous fossil assemblages at different localities throughout a region, using the rule that species that occur together most often as fossils are most likely to have lived together. Because two fossil localities are likely to vary in composition even if drawn from the same parental biota, an increasingly full picture of regional biodiversity is gained when numbers of localities are pooled. Such studies can lead to understanding the gross diversity patterns among biotic provinces. Finally, sampling of approximately contemporaneous localities can be made at the global level, in this way providing a sample of global biodiversity for the realm(s) involved.

Because most fossil localities are not directly dated by the more accurate radiometric techniques, correlation between them can only be approximate, particularly when the fossil assemblages represent different communities or provinces. This difficulty can be mitigated by using fairly coarse time units, within which most fossil localities can be placed with some confidence. Global biodiversity studies using families and genera usually employ temporal bins of from 5 to 8 million years in duration. Some studies have used geologically based time subdivisions, such as Stages; in these cases the diversity must be normalized to the differing Stage lengths.

The various approaches to help in overcoming the incompleteness of the fossil record yield binned samples of adequate size and quality to establish general levels and trends of biodiversity within well-skeletonized groups on global scales over long periods of time. The times of major diversifications or extinctions and periods of high and low biodiversities are captured by such data. Changes in the relative dominance of different major clades are also clearly demonstrated. However, the price for binning is that fine details of diversity changes are lost. Yet preliminary studies indicate that the generic record captures the same events as the family record and indeed is more sensitive to diversity changes. Furthermore, increasingly accurate dating techniques are permitting the use of ever-shorter time intervals. These ongoing refinements have confirmed the general trends described from the coarser data, while permitting the recovery of more detailed diversity information that reflects more closely the history of species richness.

IV. MICROBIAL DIVERSITY OF THE ARCHAEAN AND PROTEROZOIC ERAS

The earliest rocks known on earth date to about 4.6 billion years ago. The earliest eras in earth history, long thought to be barren of indications of life, were named the Archaean (about 4.6–2.5 billion years ago) and the Proterozoic (about 2.5 billion–543 million years ago). The most recent era was named the Phanerozoic (visible life) because it contained a fossil record. However, microbial fossils have now been found to date well back into the Archaean. The earliest are filamentous prokaryotic microbes resembling cyanobacteria ("blue-green algae"), recovered from chert beds that are over 3.5 billion years old. There are a few other Archaean records of microbial fossils that may be from other bacterial groups, representing both benthic and planktonic realms. However, bacteria and archaea have such extremely low preservation potentials that it is impossible to reconstruct their history of richness from fossil evidence. According to molecular phylogenies, a number of major bacterial branches, equivalent at least to phyla in a hierarchical system, arose before the cyanobacteria. It appears likely that there was a rapid early diversification within the bacterial and, probably, archaeal domains that produced a highly disparate microbiota well over 3 billion years ago. Considering the way in which lineages branch during evolutionary diversifications, it is likely that the microbiota was quite rich during much of Archaean and Proterozoic times. Recorded fossil diversity is much greater in the Proterozoic, but sampling is much better in rocks of that age and the recorded fossil richness simply tends to correlate with the number of formations and samples studied. One exception is in the latest Proterozoic, when sampling is highest but when the richness of fossil microbiotas declines, suggesting a real diversity drop at that time. The fossil record of richness of the Proterozoic microbiota is given in Fig. 2.

Unicellular fossils belonging to the Eukarya are difficult to separate from those of other domains. Relatively large fossil algal ribbons (1 m long, 2 mm in diameter) about 2.1 billion years of age are likely to be the first known eukaryotes. It is possible that Eukarya evolved during the Archaean. The best Proterozoic records of unicellular Eukarya are of spheroidal plankters with organic walls, sufficiently tough to be preserved in some quantity. They increase in number, richness, and morphological complexity throughout the Proterozoic. As noted earlier, great bacterial disparity is likely to have been achieved in the Archaean, and the major richness trend may be attributed chiefly to an improving fossil record rather than to increased biodiversity.

V. PHANEROZOIC MARINE BIODIVERSITY

A. Neoproterozoic Animals

The record of metazoan fossils dates from at least 570 and possibly 600 million years ago, in the late Proterozoic (the "Neoproterozoic"), but it has proven

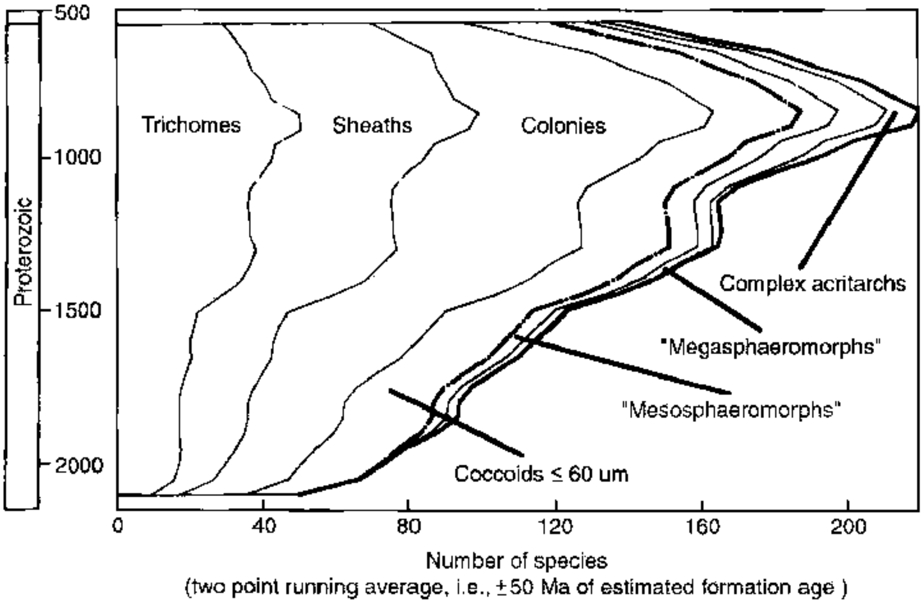


FIGURE 2 Species richness of several groups of unicellular microfossils as represented in Proterozoic sediments. The morphologic categories to the left of the heavy dashed line are presumptive Bacteria; those to the right are Eukarya. The record is unlikely to mirror the history of either richness or disparity, although the decline late in Proterozoic time may be real. (From Schopf and Klein, 1992, p. 535. Reprinted with the permission of Cambridge University Press.

to be very difficult to relate those early fossils to living groups, which first appear near the beginning of the Phanerozoic. There are two chief modes of fossilization: trace fossils, which are trails, burrows, and similar structures left in sediments by animal activities; and body fossils, which register the gross morphology of the animals. Most Proterozoic body fossils are restricted to body impressions, skeletons being absent until latest Neoproterozoic time. The early trace fossils are simple curved trails, and though traces become increasingly complex and diverse as time passes, they remain small throughout the Neoproterozoic, mostly 1 mm or less in width, though a few range to 5 mm. The early body fossil impressions (the "Vendian" or "Ediacaran" fauna) are chiefly of frondose and discoidal organisms that are large (some reach 1 m) and bear a general resemblance to cnidarians, although they differ from living forms in important structural details. Ediacaran faunas also include rare sponges and a few forms that resemble bilateral animals but that cannot be assigned to any living group. Estimating the living diversity of Neoproterozoic times from these fossils is difficult; body fossils of the trace-makers, for example, are not known, and many kinds of organisms can leave a given type of trace fossil. When body fossils appear during the Cambrian (see the following section), traces become larger (some are centimeters wide), much more varied, and more common. Therefore the best guess for Neoproterozoic animal diversity is that it was lower than that of the Cambrian, beginning sometime before 570 million years ago (Ma), possibly tens to hundreds of millions of years earlier. This guess is likely to be correct as far as body plan disparity is concerned, but it remains possible that Neoproterozoic seas supported a rather high richness of species that had such low preservation potential that they have not yet come to light.

B. The Cambrian Explosion Establishes Metazoan Disparity

About 543 Ma there was a notable increase in the complexity, sizes, and abundances of trace fossils, one of which (*Triptichnus*) marks the basal Cambrian boundary by international agreement. Rare, minute, mineralized skeletons of uncertain affinities had appeared just prior to the first appearance of *Triptichnus*, and continued to appear in increasing kinds and numbers, along with an increasing trace diversity, in younger rocks. At about 530 Ma the earliest undoubted skeletons of living animal groups appeared and, during the next 10 million years or so, 11 animal phyla made their first appearances, a period known as the "Cambrian explosion" because of the geologically sudden appearance of diverse, abundant skeletal remains. Invertebrate skeletons are then continuously present for the remainder of geologic time. Among the body plans to appear during the explosion are complex representatives of all major animal clades, including annelids, arthropods, and chordates (Fig. 3). Judging from the phylogenetic tree of animals, all living phyla had originated by the close of this interval, although some living phyla with low preservation potentials are still completely unknown as fossils.

Many of the explosion animals include features that are not present in their living allies, and thus extend the known morphological ranges of their clades. Some of these forms have such distinctive body plans that they are considered to be phyla of their own, whereas others can be nested within living phyla but not within living classes. These forms indicate that the disparity among the body plans of Early Cambrian animals was at least as great as it is today. For example, the relatively few Cambrian arthropod taxa are just as morphologically disparate as are living arthropods, which have had the benefit of over half a billion years of subsequent evolution and consist of million of species today. Thus disparity and richness are evolutionarily decoupled. The Early Cambrian burst of disparity in the fossil record suggests that there was relatively rapid evolution of the developmental regulatory gene systems at that time.

C. Phanerozoic Diversity Patterns in the Marine Environment

1. Standing Diversity of Marine Invertebrates

The combination of completeness, disparity, richness, and duration of the marine fossil record is superior to that of other realms, especially for marine invertebrates, and thus is best suited as an introduction to major diversity patterns. Figures 4 and 5 depict the overall pattern of family richness for marine animals, almost entirely shallow-water (continental shelf depth) forms. It is likely that the range of body plan disparity declined somewhat after the Cambrian, and probably showed a significant drop and then a recovery following the Permian–Triassic extinction, but there are no data that incorporate the entire marine fauna. However, Jack Sepkoski has complied an excellent data set on family richness, which has certainly varied significantly.

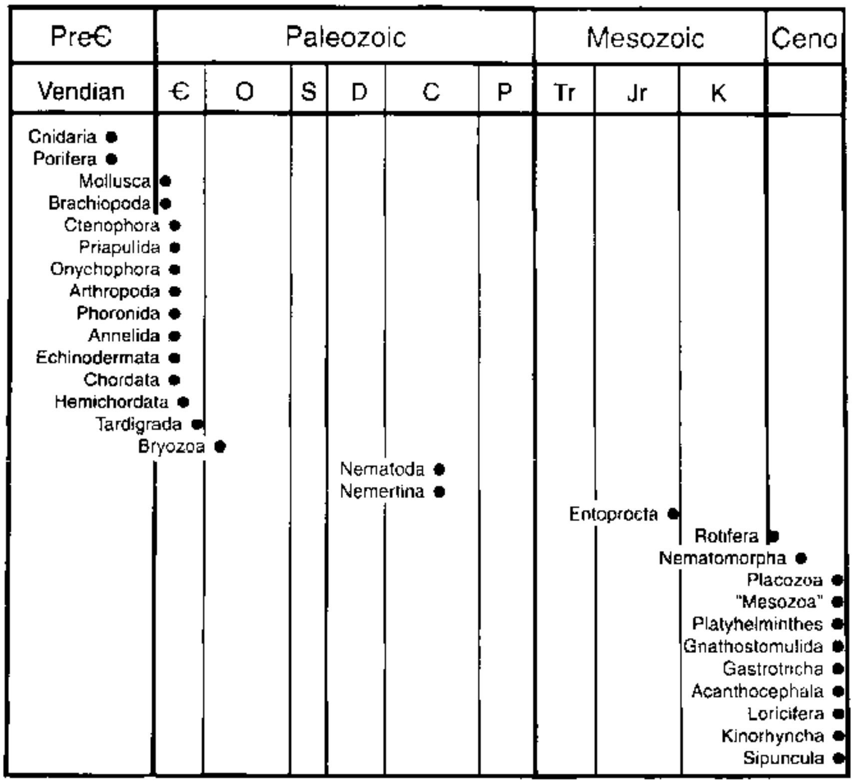


FIGURE 3 The earliest appearance of body fossils of living phyla. The earliest records of small, soft-bodied forms with low fossilization potentials are scattered through time, being found at localities with unusual preservation, such as in amber, or having no fossil records at all. The only phylum with well-mineralized skeletons that does not appear in the Cambrian is the Bryozoa, which, however, has numbers of nonmineralized forms today and may have had no mineralized representatives during the Cambrian. €, Cambrian; O, Ordovician; S, Silurian; D, Devonian; C, Carboniferous; P, Permian; Tr, Triassic; Jr, Jurassic, and K, Cretaceous. (Modified from Valentine *et al.*, 1999.)

During the Cambrian, family richness rose abruptly and seemed to be reaching a plateau as the period ended. However, the Ordovician witnessed a dramatic increase in family richness, which reached a plateau near 400 families by the end of the period (see Fig. 4). This richness plateau lasted for about 300 million years, during which there were two major extinctions that lowered family richness for a relatively brief time, followed by recoveries to the plateau level until the close of the Permian period. At that time, the plateau was abruptly terminated at about 250 Ma by the greatest extinction on record, near the Permo–Triassic (P/T) boundary. About half of the marine families were lost. From that low point during the depths of the extinction, family richness has climbed at a rate quite similar to the Cambro–Ordovician rate, and for a far longer time, peaking at the present. This last richness rise is interrupted twice by major extinctions, the last of which, near 65 Ma, is the famous Cretaceous–Tertiary (K/T) extinction that eliminated the dinosaurs. Despite these interruptions, family diversity today far exceeds that of the long Paleozoic plateau.

Perhaps as few as one-third of living marine families have durable skeletons; the rest have very poor preservabilities. Nevertheless, there is evidence to suggest that the major trends in the fossil record are real, and that the ratio of easily to poorly preservable forms has been similar enough, at least following the Cambrian explosion, that the general magnitudes of family diversities can be reconstructed from the preserved fossils. For example, the diversity patterns of easily and poorly preservable taxa are similar throughout the shallow ocean today, so far as we can tell. Further, the ratio of poorly preserved to easily preserved families that are found (see Fig. 5) holds reasonably steady through the Phanerozoic, increasing

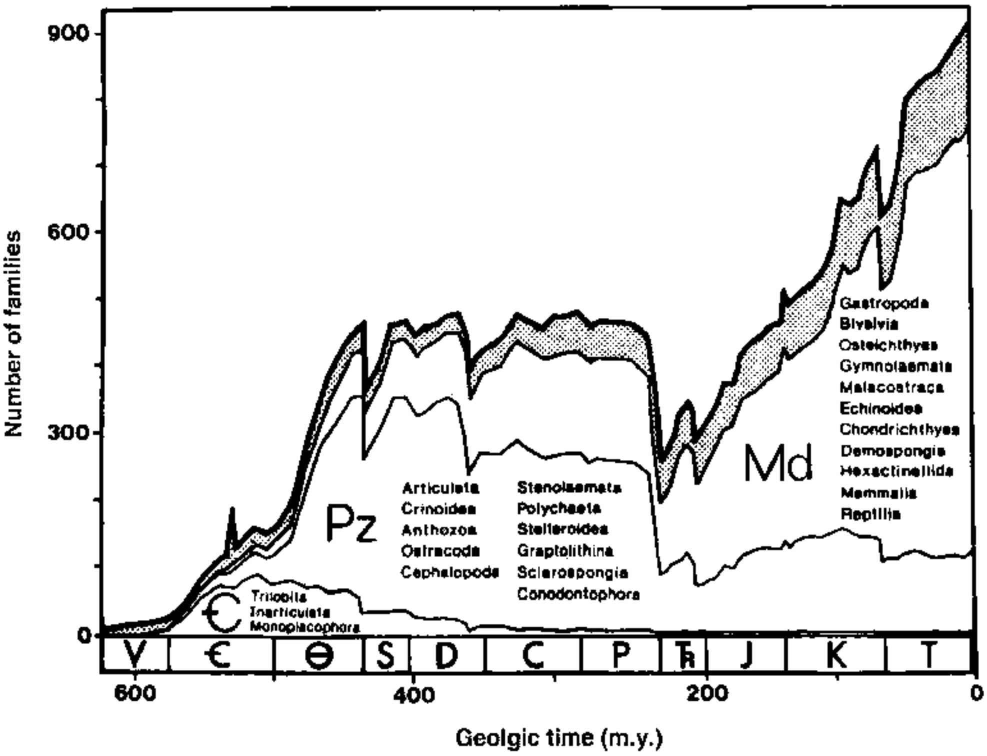


FIGURE 4 Marine animal family richness known from the fossil record. Geologic period abbreviations are as in Fig. 3. Major faunas are labeled Є, (Cambrian Fauna), Pz (Paleozoic Fauna), and Md (Modern Fauna). Shaded area at top of curve indicates animals that lack durable skeletons. (From, J. J. Sepkoski and A. Miller, in Valentine, 1985. Copyright © 1985 by PUP, Reprinted by Permission of Princeton University Press.

slightly toward the present, suggesting that the record is sampling a fairly consistent fraction of the poorly preserved clades, and the ratio improves slightly in younger sediments as would be expected. Also, the presence of a long richness plateau during the Paleozoic is unlikely to be a matter of chance, and suggests some stability in sampling success.

2. Standing Diversity of Marine Chordates

The phylum Chordata includes an invertebrate class, the Cephalochordata, that is known from the Early Cambrian and is represented today by amphioxus, (*Branchiostoma* spp.), but this class is not known to ever have been diverse. There are primitive groups within the class Vertebrata, such as lampreys and hagfish, that in fact lack vertebral columns. A group of marine vertebrates at about this level of organization, known as the Conodonta from toothlike feeding structures, appeared during the Middle Cambrian and left a significant Paleozoic fossil record. Conodonta became extinct during the Triassic. Nonconodont jawless fishes (Agnatha) appeared in the late Upper Cambrian and radiated into four distinctive, disparate clades during the Silurian; many agnathans had bony armor that has provided most of their fossil record, but the agnathans seem not to have been rich in species. They were joined by jawed fishes by Devonian time. The jawed forms radiated to produce several disparate types—arthrodires and placoderms, with bony head shields and plated armor, and several chondrichthian groups (sharks and rays), which were the richest of fish clades in Late Paleozoic seas. Also arising at least by Early Devonian time was a clade, the osteichthyes or bony fishes, which was of secondary importance during the Paleozoic but came to be the most richly diverse vertebrates in the sea during the Mesozoic and quite dominate marine vertebrate faunas today. Thus the history of marine fishes can be summed up as beginning modestly in the Cambrian and finally achieving significant disparity during Silurian and perhaps Early Devonian radiations that established most major groups, which then waxed and waned in richness, with bony fishes finally coming to dominance. Reptiles entered the sea during the earliest Mesozoic and mammals early in

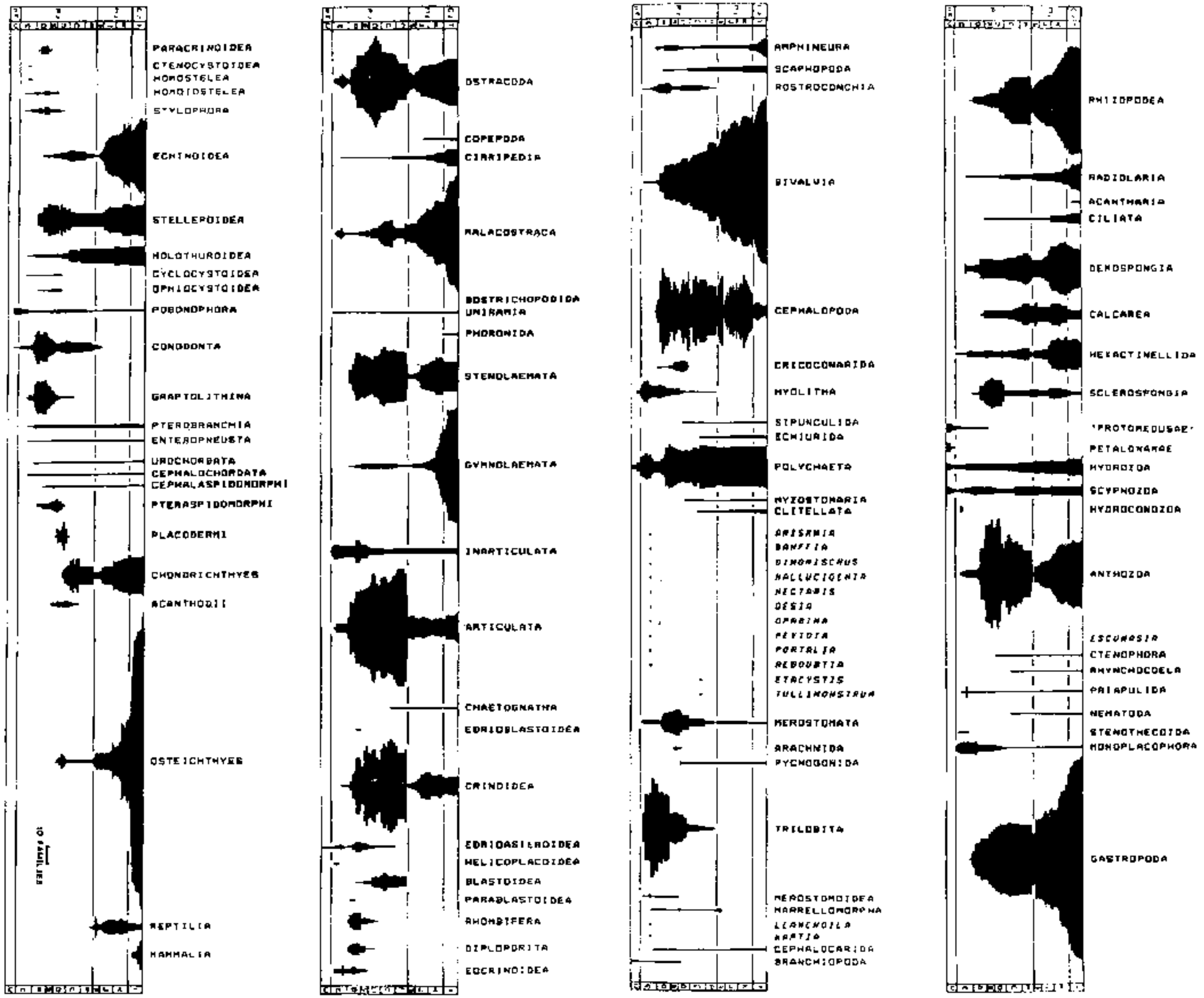


FIGURE 5 Phanerozoic diversity in the marine fossil record, showing animal family richness by phylum or class. Geologic periods are as in Fig. 3. (From J. J. Sepkoski and M. Hulver, in Valentine, 1985. Copyright @ 1985 by PUP. Reprinted by permission of Princeton University Press.)

the Cenozoic, but neither group became very rich in marine species.

3. Taxonomic Turnover

Relative clade richnesses vary considerably across Phanerozoic time. In Fig. 4, the changing assemblages of dominant clades are classed into three major faunas. The Cambrian Fauna (labeled C), composed dominantly of trilobites (Arthropoda), does not diversify during the Ordovician and is overtaken by the expanding clades of the Paleozoic Fauna (labeled Pz), dominated by articulates (Brachiopoda) and crinoids (Echinodermata). The Paleozoic Fauna is decimated by the P/T extinction and does not increase significantly thereafter, and is overtaken by the diversifying Modern Fauna (labeled Md), which is dominated by mollusks and fishes. Elements of all of the faunas are present throughout the Phanerozoic.

Figure 5 depicts the great variety of family richness histories recorded for marine classes and phyla; the dominance of the characterizing clades during each of the three faunas is clearly evident. There are some general features that link the diversity patterns to evolutionary rates that seem to be inherent in the clades. The dominant clades of the Cambrian Fauna have very high evolutionary turnover rates—high rates of both diversification and extinction. The dominant clades of the Paleozoic Fauna have moderately high turnover rates, and those of the Modern Fauna have the lowest turnover rates of the dominant clades. As a result, the extinction rate for all families combined, excluding the major extinction events, declines steadily throughout the Phanerozoic. Evidently, a strategy of being resistant to extinction is superior to having a high speciation rate, perhaps because slight inclement environmental fluctuations can force a high-turnover lineage too close to extinction to recover. Note that the presently dominant bivalve and gastropod mollusks, slow-turnover clades, increase their relative representations inexorably throughout the Phanerozoic, and are not greatly affected even by the P/T extinction. A few clades that have fair preservation potentials can be shown to have survived for hundreds of million of years despite having very low diversities; the inarticulate brachipods and the scaphopod mollusks are classic examples (see Fig. 5). These are low-turnover clades; if they had high turnovers they would presumably have been swept away by even a small extinction event. Both speciation and extinction rates are probably tied to population parameters that of course evolved without respect to clade longevity or dominance, which in a sense are merely side effects, though they profoundly influence the composition of the world's biotas.

VI. PHANEROZOIC TERRESTRIAL BIODIVERSITY

A. Plant Diversity

"Land plants" include two major groupings: forms such as mosses and liverworts, collectively termed bryophytes, which lack a water-transporting system; and the tracheophytes, which possess vascular tissues that conduct water vertically and permit significant upward growth. It seems likely that tracheophytes arose from among the mosses, perhaps more than once. The earliest fossil record of plants on land may be furnished by spores of Middle Ordovician age that are possible bryophytes; tracheophytes probably date from Early Silurian, but their vegetative elements first appear in the Middle Silurian. From then through the Devonian, a plexus of early plant types diversified, and from this plexus the modern groups of lycopods, horsetails, ferns, and gymnosperms evolved by latest Devonian time; there were forests before the close of the Devonian. This early burst of disparity is similar to that found among animals, but the richest of living plant groups, the angiosperms or flowering plants, appear much later, arising perhaps in the Late Jurassic but spreading during the Late Cretaceous and dominating Cenozoic plant associations. There is no animal analog for the late origination of a dominant major taxon, except for the class Mammalia, also terrestrial of course. However, the disparity indicated in the pollen record of angiosperms has been investigated, and it shows an early burst of disparity increase in the Early Cretaceous that gradually tapers off to a plateau by late in the Cretaceous. Evidently angiosperm "body" plans were evolved early in the history of the group and the rise of species richness followed later, the same pattern that is found within many marine invertebrate groups and in numbers of vertebrate groups. Figure 6 indicates the relative richness of major land plant groups through time as judged from their fossil records.

B. Animal Diversity

1. Invertebrates

Animals came ashore with the earliest plants, so far as can be told. It is quite possible that early terrestrial animals fed on bacteria, fungi, or protistans and were ashore before tracheophytes; early arthropod burrows are reported in Ordovician soils. However, the earliest undisputed terrestrial animal remains, of arthropods, are from the Late Silurian. Because these fossils are predators (primitive centipedes and an arachnid), there

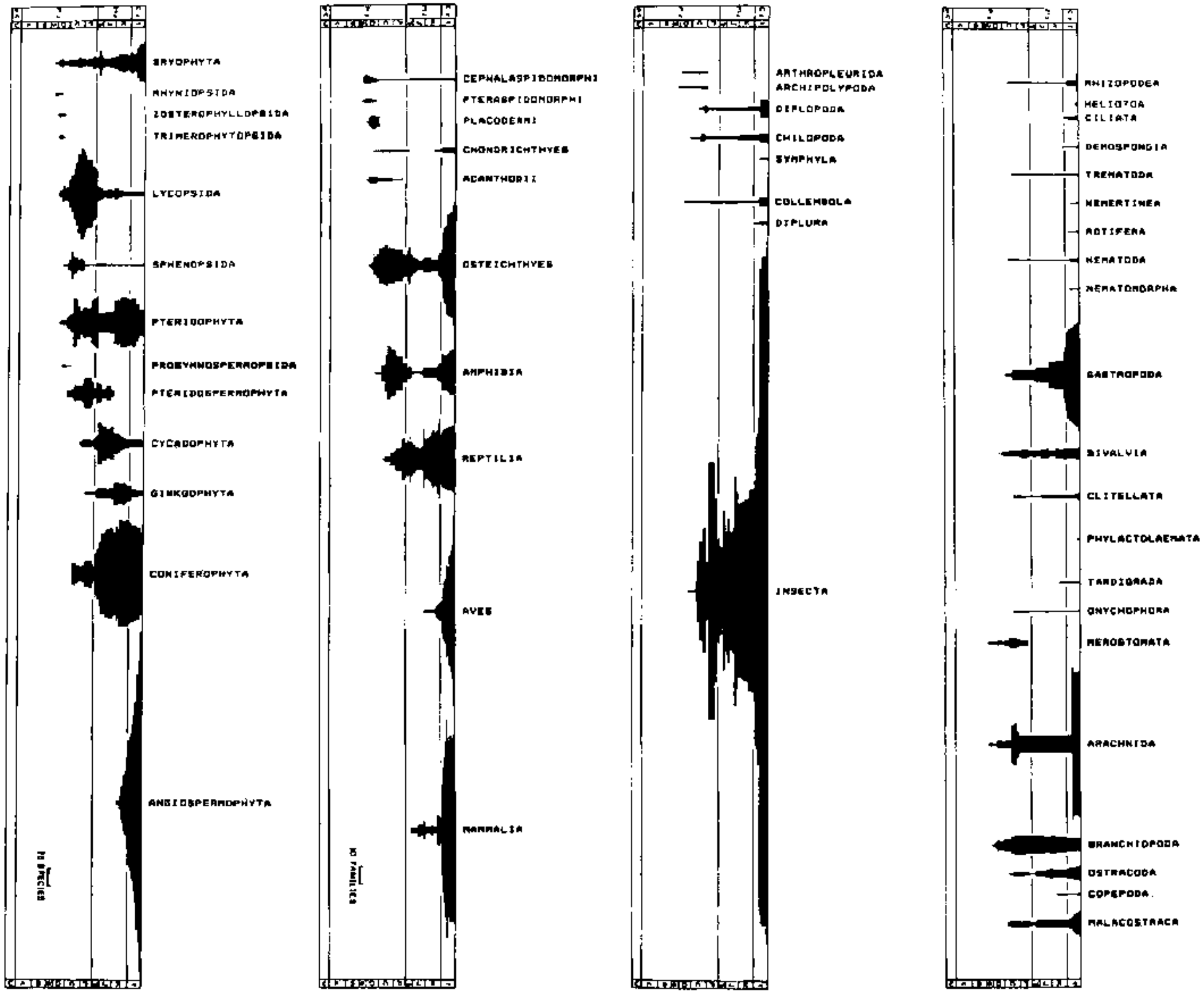


FIGURE 6 Phanerozoic diversity in the terrestrial fossil record, showing plant and animal family richness by phylum, class, or other higher taxon. Geologic periods are as in Fig. 3. (From J. J. Sepkoski and M. Hulver, in Valentine, 1985. Copyright © 1985 by PUP. Reprinted by permission of Princeton University Press.)

must have been by that time an extensive fauna of primary consumers and a fairly complex ecosystem, evolved in previous epochs. There is little evidence to judge how quickly and to what levels terrestrial invertebrate diversity rose during those times.

Insects evidently arose in the Early Devonian, and they have a relatively rich fossil record for a terrestrial group, being richer than the vertebrate record. The same species proportions found among living orders are also found among Tertiary fossils, suggesting that the insect record, while certainly incomplete, is not heavily biased taxonomically, and that at least their major fossil trends probably reflect actual paleobiological trends. During their earliest, Devonian history, insects are spottily represented; most of them are predators, whereas the primary consumers seem to have been chiefly detritus feeders; little herbivory is found. Insects underwent a significant expansion in the Carboniferous, and species richness rose to a high just before the major Permo–Triassic extinction, which affected insects significantly (see Fig. 6). However, the insects rebounded, rising steadily to their present overwhelming richness. Probably the rise was more evenly distributed in time than indicated in Fig. 6, for there are some spectacularly rich insect faunal localities in the Late Tertiary, while equivalent faunas are not yet known from earlier deposits. As with marine invertebrates, significant morphological disparity was reached very early in insect history, preceding the greatest rises in richness. Though insect richness seems closely tied with the richness of the angiosperm flora, insect disparity, even including the evolution of the array of mouthparts and feeding structures present today, was largely accomplished well before the origin of angiosperms.

2. Chordates

All the primitive groups of jawed fishes are found in terrestrial aquatic deposits during the Middle Paleozoic, but on land as in the sea it was the bony fishes (Osteichthyes) that radiated to produce rich faunas that dominated in fresh waters, from the Devonian onward. The richness of terrestrial fossil fish faunas declined significantly in the Mesozoic, perhaps partly owing to a poor aquatic fossil record from that Era.

Tetrapods first appear in the Devonian, possibly Early Devonian, in effect as part of the general radiation of fishes, but with limbs that eventually permitted locomotion on land and led to the evolution of the Amphibia. The fossil record of this group is very sparse, but it is clear that the amphibians diversified extensively to reach a level of family richness comparable to today's by the close of Carboniferous time. The early amphibians were rather different from living forms, however, and included large animals, which were eventually replaced ecologically by reptiles. Three important groups arose from amphibian stocks, one leading to reptiles, one to mammals, and one to dinosaurs and birds. The early histories of these groups are so poorly known that their patterns of diversification cannot be accurately reconstructed. It is possible that both reptiles and dinosaurs gradually increased in richness and diversity through most of the Mesozoic era. The richness of these clades was reduced drastically during the extinction at the close of the Cretaceous.

While mammals must have diversified as well during the Mesozoic, it is in the early Cenozoic that their fossil record is marked by a burst of new appearances and, by about 10 million years after the start of the Cenozoic, all of the 17 modern orders of placental mammals had appeared. Only two of these orders are known from the Cretaceous, and while it is likely that most trace their roots to a few tens of millions of years earlier, mammals were clearly a minor though probably not unimportant part of the tetrapod fauna before the Cretaceous–Tertiary (K/T) extinction. It is usually assumed that mammals replaced dinosaurs, which became extinct at the K/T boundary, in ecological roles that permitted large body sizes. The only dinosaur descendants today are birds, which have a poor fossil record. Modern bird orders are unknown before the Cenozoic, suggesting that there was a significant K/T extinction of birds, leading to the radiation that produced modern types. There must indeed have been a significant diversification of birds during the Cenozoic, but as the roots of the modern lineages are uncertain and the bird record is poor, the pattern of bird diversity increase has not been established.

VII. PRINCIPAL BIODIVERSITY FACTORS

A. Plate Tectonics and Global Heterogeneity

The geographic pattern of land and sea is constantly changing due to the processes of plate tectonics, which create and destroy the crust of the earth. At present the crust is divided into about 6 major and 14 minor crustal plates. New crust is added at one plate margin, a rift marked by deep-sea ridges, from sources of molten rock in the earth's interior. At the opposite plate margin, the crust plunges back into the interior, its descent marked by deep-sea trenches or "subduction zones"; this crust then remelts. Continents or islands on a crustal plate therefore move across the earth's surface, from the constructional margin toward the destructional margin, as if on a conveyor belt. Continents separated by a rift will diverge as they are carried apart on their respective plates, and continents will fragment if a rift cuts across them. On the other hand, continents separated by a subduction zone will converge, and if they meet they can weld together. The rates of movement of the plates are only several centimeters per year. Nonetheless, because plate tectonic processes have been incessant since well before the origins of animals and plants, the continents and oceans have changed their numbers, sizes, and geographic configurations over the many millions of years since those organisms first evolved.

As continents drift across the earth's surface they enter new climatic zones, and as oceans enlarge or shrink or change their shapes, the oceanic circulation changes, with warm or cool currents directed to different regions. The biota of any given area will evolve to adjust to the changing conditions, and may become enriched, perhaps if moving into the tropics, or depauperate, perhaps if moving into high latitudes. At the same time, migration routes are opened or closed by the shifting geography, permitting invasions of biotas into some regions while isolating the biotas of other regions. The biodiversity patterns both on land and in the sea reflect the environmental patterns created by plate tectonic processes.

In addition to its effects on rearranging regional biodiversity patterns, changes in global geography can significantly affect global biodiversity levels. The most impressive example is in the rise of biodiversity during the last 200 million years or so, from the Middle Mesozoic era through the Cenozic era. At the beginning of the Mesozoic, the continents were welded together into a supercontinent, called Pangaea. Although Pangaea was certainly environmentally heterogeneous, with large climatic and topographic variations, the climatic zones tended to be occupied by widespread biotas that could spread across what are now separate continents. In the seas, the shelf encircled the single supercontinent in an essentially uninterrupted band. While the distribution of species was limited by climatic zonation and perhaps regional habitat peculiarities, even shallow-water species tended to be relatively widespread and of course were largely confined to the one shelf, though perhaps some were present on oceanic islands, of which little is known. As the Mesozoic passed, however, the continents were traversed by rifts and began to break up into smaller landmasses, between which were arms of the sea that finally widened into oceanic expanses. By the close of Early Cenozoic time, all of the present continents that are now separated had broken from each other.

Species that could once range widely across Pangaea became isolated on separate continents. As the continents dispersed into different climatic zones and developed different topographies, their terrestrial biotas each evolved distinctively in response to local conditions and events. For example, North America, Eurasia, and Australia each developed distinctive temperate biotas, and the tropics of the Americas became distinct from the Old World tropics and those of the western Pacific. In the sea, the growing geographic differentiation was even more profound, as new shelves appeared on each side of the new oceans, significantly raising the number of isolated shelf segments. As on land, each continent evolved a distinctive biota in response to the unique set of environmental conditions and events that developed in each region. The heterogeneity of world environments thus increased greatly during this period, on a global scale. These plate tectonic effects are responsible for a significant part of the important rise in standing global marine diversities during the Mesozoic and Cenozoic as represented at the family level in Fig. 4; an accompanying rise occurred in the continental biota.

B. Climate Change

Climate is another major factor governing global biodiversity. For example, earth underwent a general cooling trend during the Cenozoic, although there were shorter warming and cooling events superimposed on the generally falling temperature curve. The temperature change was greatest in high latitudes, thus increasing the temperature differential between the poles and the equator. The margins of the tropics were shifted toward the equator, the tropics became compressed, and cooler climates came into existence in high latitudes, shifting equatorward themselves as cooling continued, to be replaced poleward by still cooler climatic regimes.

In the sea, the increasing temperature gradient made for a markedly increased provincialism. Tropical species whose biology was related to climate became progressively more restricted to lower latitudes, with some occasional, relatively short-lived reversals in which species tolerant of lower temperatures arose and diversified in the more poleward regions. Because the major rifting that produced the Cenozoic continental pattern was primarily in a north–south direction, most major continental shelves trend north–south in the Cenozoic, as they do today, and so there was a rise in species richness on each shelf as the latitudinal thermal gradient increased, multiplying the effect on global marine biodiversity by the number of isolated shelves. The Cenozoic global marine rise in family richness shown in Fig. 4 is compounded partly of the increase in the number of isolated shelves following continental rifting and partly of the increased provincialism on each shelf created by an increasing thermal gradient; together they have formed a powerful engine of global species enrichment.

In terrestrial environments, family richness also increased during those Cenozoic events, and for the same reasons. The continental expanses provided for a three-dimensional array of habitats (including high mountains and basins) unlike the rather two-dimensional array along the narrow, shallow continental shelves, and therefore the biotic response has been correspondingly more complicated in terrestrial than in marine environments. Not only did high-latitude climate zones open up, but high-altitude regions became cooler, increasing the contrasts between mountain and lowland environments. A lowering of rainfall in some continental interiors produced semiarid plains and steppes. The biota responded to these increases in environmental heterogeneity by producing, for example, alpine and grassland plants and animals, thus enriching both the flora and fauna. The species richness of angiosperms among plants and of insects among animals profited the most from these events.

During the Late Cenozoic, as cooling has continued in high latitudes, massive ice accumulations on a subcontinental scale have produced a series of glacial ages interrupted by warmer interglacial periods, most pronounced during the Pleistocene, which began about 1.6 million years ago. Some regional extinctions occurred during the early onset of cold periods, but in general there has been little extinction, and thus no significant lowering of global biodiversity, associated with these climatic swings. Instead, the biotas of both land and sea have tended to migrate with their climatic zones, southward during glaciations and northward during interglacial times. Thus regional diversities have changed as species have migrated, but overall diversity has not been significantly affected. Indeed, the Recent biosphere may be the richest in species during the entire history of animals and plants, at least prior to the deleterious influences of human activities on biodiversity.

C. Extinction Events

Changing environments have often provided opportunities for the origin of species, some of which have clearly led to major morphological novelties and enhanced morphological disparity. At the same time, changing conditions have led to the extinction of species, which has occurred more or less continually over Phanerozoic time. As noted earlier, extinction rates vary among taxa, whose histories tend to reflect their rates of turnover. The species richness of a taxon results from an interaction between its rates of speciation and extinction. Thus when there has been some unusually profound disturbance to the global environment, and extinction rates have become unusually high, species richness has fallen dramatically in events called mass extinctions. The most drastic such event, recorded in marine fossils, was the Permian–Triassic (P/T) extinction about 250 million years ago, at the close of the Paleozoic era. As many as 90% of marine species may have become extinct across the P/T boundary (see Figs. 5 and 6). This extinction was highly differential; the groups that did best in surviving the extinction and in rediversifying were those that dominate the marine fauna today. Disparity as well as richness was affected; some entire classes and orders of marine invertebrates disappeared. Brachiopods, crinoids, and other groups that were dominant before the extinction were reduced to minor roles. The causes of this extinction, and the extent of its effects on the terrestrial biota, are not yet clear.

There have been a number of other mass extinctions (five major events in all) and numerous regional extinctions that were severe locally. The most famous of the mass extinctions, the K/T, occurred about 65 million years ago at the close of the Mesozoic era, although it did not have as profound an effect on the marine environment as the P/T event. The K/T extinction strongly affected terrestrial as well as marine life. It probably resulted from the impact of an extraterrestrial object, such as a meteor, with the earth.

The "background" extinctions that account for the general turnover of taxa are accompanied by "background" speciations, so that species richness levels do not change much during normal turnover and tend to be significantly affected by background effects only during trends that last tens of millions of years or more. However, the more abrupt, severe regional and mass extinctions are not immediately compensated by correspondingly massive speciations, and indeed the recovery of species richness usually requires millions of years following a mass extinction. For phyla, classes, or orders that are lost to extinction, there is no recovery at all, and the unique gene regulatory systems involved in their architectures are gone forever.

Clearly, biodiversity data from the fossil record support the notion that species richness is largely associated with environmental heterogeneity, and thus the preservation of species diversity can be promoted by preserving habitat diversity. There are entire animal phyla that are represented today by very few species: the phylum Phoronida by only 12 or so, and the phylum Priapulida by perhaps 17, for example. These phyla harbor uniquely organized genomes that have been present for over half a billion years of earth history, and that are largely unstudied. Their preservation should be a priority, for they represent major elements in the disparity of the present biosphere.

See Also the Following Articles

ARCHAEA, ORIGIN OF • BACTERIAL BIODIVERSITY • BIODIVERSITY, ORIGIN OF • BIOGEOGRAPHY, OVERVIEW • CLADOGENESIS • EUKARYOTES, ORIGIN OF • GENES, DESCRIPTION OF • MICROBIAL BIODIVERSITY • TAXONOMY, METHODS OF

Bibliography

Bengtson, S. (ed.) (1994). *Early Life on Earth.* Columbia University Press, New York.

Foote, M., and Sepkoski, J. J. Jr. (1999). Absolute measures of the completeness of the fossil record. *Nature* 398, 415.

Pace, N. R. (1997). A molecular view of microbial diversity and the biosphere. *Science* 276, 734.

Schopf, J. W., and Klein, C. (eds.) (1992). *The Proterozoic Biosphere.* Cambridge University Press, Cambridge, United Kingdom.

Valentine, J. W. (ed.). (1985). *Phanerozoic Diversity Patterns.* Princeton University Press, Princeton, New Jersey.

Valentine, J. W., Jablonski, D., and Erwin, D. H. (1999). Fossils, molecules and embryos: New perspectives on the Cambrian explosion. *Development* **126**, 851.

GOVERNMENT LEGISLATION AND REGULATION

Kathryn A. Saterson
Brandywine Conservancy

GLOSSARY

case law Interpretations of the U.S. Constitution, statutes, and regulations provided by the judicial branches of U.S. federal or state government. Such interpretations are provided when two or more parties disagree as to the meaning of a law in a specific context and bring it to the courts to decide.

executive order A directive from the U.S. president, or a state governor, specifying actions by government officials and agencies. State and federal executive orders often describe how to administer a provision of a statute, treaty, or the constitution and usually have the force of law.

regulations Rules and administrative codes required by statutes and issued by local, state, and federal government agencies. Regulations have the force of law since they are adopted under the authority granted by statutes.

statutes or legislation Laws passed by the legislative branch of U.S. federal or state governments. The U.S. Congress passes federal laws and state legislatures pass state laws. Local laws are usually called municipal ordinances.

treaties Multilateral agreements between or among nations. International treaties, often called "conventions," are entered into in the United States by the president but must be ratified by the Senate. Treaties supersede federal, state, and local laws that might have contradictory goals.

BIOLOGICAL DIVERSITY, at the genetic, species, and ecosystem levels, is influenced by legislation and regulation that is initiated and enforced at local, national, and international levels. Although a great many laws and regulations governing human activities have indirect impacts on biodiversity, this article will summarize only major laws and regulations that directly affect biodiversity by influencing how genes, species, and ecosystems are used and conserved. Although the legal measures taken by each nation have an impact on global biodiversity, for brevity this article will focus primarily on U.S. laws.

I. INTRODUCTION

For centuries, customary or traditional laws have governed the use of biological resources throughout the world. The systems of customary law in traditional cultures often dictate limits on how much of a specific biological resource can be used and in what seasons. In many societies an individual's rights to biodiversity are not always linked to land tenure. As traditional societies are becoming more integrated into national and global markets and political systems, traditional knowledge and laws are being lost. As human population levels increase in all societies, voluntary actions governed by cultural practices are sometimes insufficient to protect biodiversity, leading to an increasing reliance on the creation and enforcement of government legislation and regulation to protect biodiversity.

A variety of motivations have led to creation of legislation intended to help conserve biodiversity. Early legislation was primarily aimed at protecting biodiversity as an economic resource for humans (e.g., for hunting or agriculture). Recent laws, and amendments to earlier laws, have included other reasons, such as the value of biological diversity to overall ecosystem health and the intrinsic value of species. Although U.S. national parks were originally created primarily for scenic and recreational values, the 1916 National Park Act identified conservation of wildlife as an important goal of parks.

Laws and regulations can aim to influence actions undertaken by a range of actors (government agencies, private corporations, and individual citizens) and to affect a range of resource types (animals, plants, water, land, and specific habitats) with differing ownership and resource control (public and private). The success and evolution of legislation and regulation in the United States are in part due to the actions of nongovernmental organizations that lobby for new or amended legislation, oversee how agencies implement legislation, and use the judicial system to help ensure that laws are enforced and interpreted appropriately.

All laws and regulations in the United States are based on rights found in the U.S. Constitution. The federal government derives its authority to regulate wildlife from its constitutional authority to regulate commerce, to protect its own property (federal lands), and to make treaties. When federal laws conflict with state laws, the federal laws generally take precedence.

Many advocates of the right to private property, also supported by the U.S. Constitution, view legislation to protect biodiversity as sometimes conflicting with their right to private property. The Fifth Amendment to the Constitution requires that when private property is taken for public use (whether to benefit public safety, create roads, protect wildlife, or for any other purpose) there must be "just compensation." Most state and federal case law indicates that restrictions on killing wildlife are not a sufficient taking of private property to require compensation by the government (Bean and Rowland, 1997).

Despite the large number of environmental laws and regulations that now exist, and the many positive impacts that they have had, these positive steps have been insufficient to prevent the loss of biodiversity. Some argue that federal and state governments give greater weight to protecting current economic growth and private property rights than to protecting biodiversity because of a lack of constitutional authority to protect future generations. Some conservationists advocate an amendment to the U.S. Constitution that obligates the government to protect biodiversity for future generations. Many other nations and at least 12 of the United States require conservation of nature in their constitutions. For example, the Constitution of the Commonwealth of Pennsylvania (Art 1, Sec 27) states that people have a right to preservation of natural values and that "natural resources are the common property of all people, including generations yet to come. As trustee of these resources, the Commonwealth shall conserve and maintain them for the benefit of all the people."

One of the controversial issues confronting the agencies and courts in implementing all the legislation described in this article is determining how the laws apply to American Indian reservations, which are covered under separate treaties, and to traditional subsistence use of biodiversity by Alaskan natives. Balancing the needs of people who depend on biological resources for food and shelter with the need to protect wildlife is an issue with both legal and ethical challenges.

This article will summarize only legislation and regulations that aim to directly conserve biodiversity. Table 1 presents a summary of selected laws intended to help conserve biodiversity. Many other environmental laws have indirect positive impacts on biodiversity, for example, by controlling air pollution (Clean Air Act), water pollution (Clean Water Act), hazardous waste (Resource Conservation and Recovery Act), global climate (Global Climate Change Convention), trade (General Agreement on Tariffs and Trade) or by creating conservation incentives through tax policies (Internal Revenue Code). This article will also not address the federal, state, and local programs (often resulting from legislation) that help to conserve biodiversity.

II. INTERNATIONAL TREATIES

International agreements are important for conserving biodiversity because many species have very large natural ranges or migrate between countries, most habitats do not follow national boundaries, and the earth is increasingly globalized in trade and movements of people. Under international treaties, the ratifying countries

TABLE I

Selected Legislation Influencing Biological Diversity

Law (year)[a]	Purpose
Multilateral international agreements	
International Convention for the Regulation of Whaling (1946)	Designates protected whale species and hunting limits
Convention on International Trade in Endangered Species of Wild Fauna and Flora (1973)	Prohibits and regulates international trade of specified plants and animals
Convention on Biological Diversity (CBD, 1993)	Encourages nations to conserve, use, and share the benefits from conservation of genetic, species, and ecosystem levels of biodiversity; the United States is not yet a party to CBD
U.S. federal laws	
Lacey Act (1990)	Prohibits commerce in wildlife injurious to agriculture and in wild animals and some plants taken in violation of U.S. and foreign laws.
National Park Service Act (1916)	Requires conservation of scenery, natural and historic objects, and wildlife in national parks and other protected areas for public benefit
Migratory Bird Treaty Act (1918)	Implements treaties with Canada, Mexico, Japan, and the Soviet Union to establish hunting seasons and protect habitat for migratory birds
Bald Eagle Protection Act (1940)	Prohibits killing or possessing bald and golden eagles, their nests, or parts
U.S. Agricultural Research and Marketing Act (1946)	Requires collecting, preserving, and disseminating genetic material important to U.S. agriculture
Foreign Assistance Act (1962)	Requires that U.S. foreign aid to developing nations helps to protect biodiversity and tropical forests, and that development projects generally avoid adverse impacts on biodiversity
National Environmental Policy Act (1969) (signed by the president in January 1970)	Requires Environmental Impact Statements for federal programmatic and site-specific actions
Marine Mammal Protection Act (1972)	Creates a moratorium on importing marine mammals and limits taking of U.S. marine mammals
Endangered Species Act (1973)	Prohibits or restricts taking of endangered or threatened species, requires federal agencies to protect their critical habitat, and regulates their trade; requires evaluation of the impact of federal projects on endangered and threatened species.
Fishery Conservation and Management Act (1976)	Requires plans from state/federal councils to manage fisheries in a 200-mile offshore exclusive economic zone to ensure a maximum sustainable yield
National Forest Management Act (1976)	Requires management plans for all national forests, including conservation of a diversity of plant and animal communities
Wild Bird Conservation Act (1992)	Limits imports of exotic birds to protect wild populations in country of origin and reduce inhumane treatment of birds
National Wildlife Refuge System Improvement Act (1997)	Requires that National Wildlife Refuge System conserve fish, wildlife, plants, and their habitat while maintaining the biological integrity, diversity, and environmental health of the system

[a] This is the year legislation was first passed. All laws shown have been amended subsequent to passage. The stated purpose may include objectives added in the subsequent amendments and/or may not include all the purposes of the law.

agree to take specific actions at the national level. Although the earliest international treaties reflected concerns that one nation might overexploit a species that many nations found economically important, recent treaties have acknowledged the importance of conserving all levels of biodiversity, whether or not there is an immediate economic value. Ratification of a treaty by each country is voluntary, and enforcement is often difficult. The United States is signatory to many international treaties relating to biodiversity, some dating back to the early 1900s.

A. International Convention for the Regulation of Whaling (1946)

The major nations then involved in whaling (including the United States, Japan, and the Soviet Union) were party to the 1946 Convention to Regulate Whaling, which prohibits killing calves, immature whales, or females accompanied by calves and encourages using as much of the killed whales as possible. The convention created the International Whaling Commission and gave it the authority to designate protected species and to establish hunting methods and seasons, size limits, and take limits. The convention had the conflicting goals of both conserving whale stocks and further developing the whaling industry. Most species of whales continued to decline dramatically until a 10-year moratorium on commercial whaling was agreed to by signatory countries in 1986. By that time, only Japan, the Soviet Union, and Norway were still whaling. The moratorium is still in effect, although it is not adhered to by Norway, and many countries continue to take whales for the "research purposes" allowed for under the convention.

B. The Convention on International Trade in Endangered Species of Wild Fauna and Flora (1973)

The Convention on International Trade in Endangered Species (CITES), which was introduced in 1973 and entered into force in 1975, prohibits international trade in endangered species by assigning each protected species to one of three lists. Appendix I lists "all species threatened with extinction which are or may be affected by trade." They cannot be traded primarily for commercial purposes, and trade in the species cannot be detrimental to the species survival. Trade in Appendix I species requires both import and export permits approved by the "management authority and scientific authority" of the nations involved (the secretary of interior in the case of the United States). The permit has to certify that the specimen was obtained legally, that it can be shipped without risk of cruel treatment or harming the health of the specimen, and that the trade is not detrimental to survival of the species.

Species listed in Appendix II are not necessarily currently threatened, but unregulated trade could jeopardize their survival. Two-thirds of the parties to CITES must agree in order to add or remove a species from Appendix I and II lists. Any nation can unilaterally add or remove a species from Appendix III if it believes that cooperation from other nations is needed to control trade in a vulnerable species under their jurisdiction. Only export permits are required for Appendix II and III species. By 1999, more than 140 countries had ratified the treaty.

C. Convention on Biological Diversity (1993)

The International Union for the Conservation of Nature began encouraging the creation of a global biodiversity treaty in the early 1980s. The United Nations Environment Program led the process of creating the convention, which was agreed to in 1992 and entered into force in 1993. By 1998, 174 countries had ratified the treaty. The United States signed the Convention on Biological Diversity (CBD) in 1993, but it has not yet been ratified by two-thirds of the Senate. Until ratification, federal executive agencies are expected to conform with the treaty to the extent possible, but it is not recognized as law in federal and state courts.

The CBD outlines objectives for the conservation, sustainable use, and equitable sharing of the benefits from biodiversity. Signatory countries are called on to develop national strategies for protection of biodiversity, establish a system of protected areas, conduct environmental assessments of projects in order to prevent adverse impacts on biodiversity, enact laws regulating conditions of access to genetic resources, and support a financial mechanism for developing countries to obtain grants to assist in meeting the terms of the treaty. The convention supports conservation of agrobiodiversity (the plants and animals that contribute to food security for humans) as well as wild biodiversity. The importance of habitat conservation and of controlling alien species is acknowledged in the treaty.

Some of the most discussed provisions of the treaty (and the basis for the U.S. reluctance to ratify it as of 1999) are those concerning access to genetic resources and the "intellectual property rights" resulting from commercial use of genetic resources. Countries with large biotechnology industries (e.g., pharmaceutical and agribusiness) supported continued open access to genetic resources, whereas countries supplying biodiversity to industry wanted to share the profits to help ensure continued protection of biodiversity. The treaty both acknowledges each nation's sovereign control of its biodiversity and encourages access to genetic resources by other contracting parties.

III. NATIONAL LEGISLATION

Most national laws discussed here are primarily concerned with actions influencing biodiversity in the United States. A few, including the Foreign Assistance Act, legislate how actions of the U.S. government relate to conservation in other countries. The U.S. states were solely responsible for conserving wildlife until states urged the federal government to pass the Lacey Act in 1900. States still retain primary responsibility for biodiversity. The role of the federal government is limited to managing federal lands; conserving economically important genetic resources; protecting migratory waterfowl, marine species, birds of prey, and endangered and threatened species; and enforcing international treaties. The following discussion describes U.S. laws supporting conservation at the genetic, species, and ecosystem levels according to their primary purpose. Such divisions are arbitrary to a degree since genes are best conserved in the organisms in which they exist, and species depend on their habitats for survival.

A. Conservation of Genetic Resources

Federal laws regarding genetic resources are primarily concerned with conservation of varieties of economically important domesticated plants and animals, with some attention to "wild genetic resources," e.g., wild relatives of plants or animals known to be economically important. For wild species, legislation is more concerned with access to genetic resources and sharing of benefits if those resources end up being economically valuable. In 1990, an amendment to the 1946 U.S. Agricultural Research and Marketing Act modified the National Genetic Resources Program in the Agricultural Research Service in the Department of Agriculture. The purpose of this program is to collect, preserve, and disseminate genetic material that is important to American agricultural production. This amendment expanded the types of genetic resources in the program to include silvicultural species, animals, and aquatic and microbiological organisms. Congress authorized creation of the National Seed Storage Laboratory in Colorado by the U.S. Department of Agriculture. Although this gene bank is world renowned for the variety of agricultural species it contains, it has been criticized for inadequately maintaining seed from wild relatives of important crops.

B. Conservation of Species

The earliest federal wildlife legislation in the United States was concerned with species, both migratory and nonmigratory, with ranges that went beyond state boundaries. Most early legislation emphasized regulation of consumptive uses of species, such as hunting and fishing. By the 1970s, Congress was aware that habitat degradation is a significant threat to biodiversity; this is reflected in the habitat conservation provisions of the Endangered Species Act and in the increase in laws concerning habitats described in Section III,C.

1. Lacey Act of 1900

The Lacey Act was created to strengthen the ability of states to protect wildlife by regulating interstate commerce in game birds and wild birds. The Lacey Act now also prohibits importing all wildlife and some plants that were taken contrary to the laws of another country. The original act also aimed to protect U.S. agricultural and horticultural interests by prohibiting the import of injurious foreign wildlife, including mongooses, English sparrows, and other birds and animals. Later amendments broadened the definition of injurious wildlife to include wild birds, wild mammals, fish, amphibians, and reptiles and their eggs, and the protected interests were also expanded to include humans, forestry, and U.S. wildlife resources. The 1981 amendments also increased the maximum penalties and lowered the standard of proof for a violation, making the Lacey Act one of the laws most frequently used by federal law enforcement officials to protect wildlife (Bean and Rowland, 1997).

2. Laws Controlling the Introduction of Nonnative Species

One of the major threats to biodiversity is the introduction of nonnative species to new habitats, intentionally or by accident. The federal authority to regulate interstate commerce led to creation of the Federal Noxious Weed Act of 1974 and the Nonindigenous Aquatic Nuisance Prevention and Control Act of 1990. The Noxious Weed Act gives the secretary of agriculture authority to regulate both foreign imports and interstate commerce in potentially harmful plants, defined as plants of foreign origin that can directly or indirectly injure agriculture, natural fish and wildlife resources, or public health. Amendments in 1990 give authority to the Fish and Wildlife Service, the National Park Service, and the Bureau of Land Management to control undesirable plants on their lands.

The accidental introduction of the economically destructive zebra mussel into U.S. waters when a ship's ballast water was released led Congress to create the Nonindigenous Aquatic Nuisance Prevention and Control Act in 1990. This act covers any plant, animal, or other "viable biological material," such as a virus, that disperses to an aquatic ecosystem in which it is not historically found. In contrast to the Lacey Act provisions, a nonindigenous species does not have to be from a foreign country. To be termed a nuisance, a nonindigenous species must threaten the abundance or diversity of native species or the ecological stability or commercial productivity of the infested waters. The act mandates creation of a task force to implement a program to prevent the introduction and dispersal of aquatic nuisance species, but it does not specify who determines which species are a nuisance (Bean and Rowland, 1997).

3. Birds

a. Migratory Bird Treaty Act of 1918

This act was passed to implement a treaty signed in 1916 between the United States and Great Britain with the aim of conserving birds that migrate between the United States and Canada. The treaty establishes closed seasons for hunting and prohibits taking of nests or eggs (except for scientific purposes) for three groups of migratory birds (game, insectivorous, and nongame—a classification that includes almost all birds). The act also implements treaties signed with Mexico in 1936, Japan in 1972, and the Soviet Union in 1976. The latter two treaties include provisions for protecting bird habitat, for example, from pollution. Numerous cases have resulted in judicial interpretations that "taking" under the treaty act does not include habitat destruction. The act was successful in supporting the 1991 requirement that migratory waterfowl not be hunted with lead shot. The courts have rejected landowner claims that the hunting prohibitions under this act are unlawful "takings" of their private property (Bean and Rowland, 1997).

b. Bald Eagle Protection Act of 1940

The Bald Eagle Protection Act prohibits killing or possessing bald eagles, their nests, or any part of the eagle. The secretary of interior could approve exceptions for scientific and educational purposes. A 1962 amendment added the same protections for golden eagles and authorized the secretary to permit taking of bald and golden eagles by Indian tribes only for religious purposes. A 1972 amendment of the act expanded the definition of prohibited taking to include poisoning.

c. Wild Bird Conservation Act of 1992

The United States is the largest importer of wild birds in the world. It is estimated that as many as 50% of captured exotic wild birds die before they even leave the country of origin, and another 15% die in transit to the United States (Bean and Rowland, 1997). The goal of the Wild Bird Conservation Act is to promote conservation of birds not indigenous to the United States by (i) limiting or prohibiting imports of nonindigenous birds in order to reduce depletion of wild populations and reduce inhumane treatment of birds in transit, (ii) ensuring that trade is biologically sustainable, and (iii) assisting in wild bird conservation programs in the country of origin. Congress found that CITES was not effective in decreasing the rate of loss of wild bird populations because many signatory exporting countries were unable to implement it adequately. The act allows the secretary of interior to declare a moratorium on importing any exotic bird listed on any CITES appendix and to permit exceptions for scientific purposes.

4. Marine Mammal Protection Act of 1972

The Marine Mammal Protection Act (MMPA) replaced individual state authorities and programs concerning marine mammals and created a moratorium on importing marine mammals to the United States and on taking of U.S. marine mammals. Taking was defined to include attempts to harass, hunt, capture, kill, and (after later amendments) feed marine mammals in the wild. Responsibility for implementing the act rests with the secretary of commerce for (i) incidental take of mammals during commercial fishing, (ii) all Cetacea (whales and porpoises), and (iii) all Pinnipedia (seals) except for walruses. The secretary of interior implements the MMPA for all other marine mammals, including manatees, polar bears, sea otters, and walruses. Like most environmental legislation, the MMPA reflects a compromise between many interest groups: commercial interests wanting to protect an economic product, scientists who valued marine species and their role in marine ecosystems, and animal welfare groups wanting to protect marine mammals for their individual and intrinsic value.

The MMPA outlines management principles that include maintaining the health and stability of marine ecosystems and the "optimum sustainable population" for all marine mammals as well as the "maximum sustainable yield" for commercially exploited species. The taking prohibitions allow many exceptions; one that received the most attention was the controversial allowance for "incidental taking" during commercial fishing operations. At the time the act was passed, more that 5 million dolphins had died accidentally by drowning in purse seining nets used to catch tuna (Bean and Rowland, 1997). Beginning in 1976, a series of regulations set quotas on how many dolphins could be killed incidentally by U.S. tuna fishing boats. In 1984, amendments sought to ensure that countries exporting tuna to the United States had kept incidental dolphin kills as low as those by the U.S. fleet; otherwise, a moratorium could be imposed on importing tuna from that country.

The 1992 International Dolphin Conservation Act required no import to the United States after 1994 of any tuna caught in association with incidental dolphin mortality (Bean and Rowland, 1997). This and similar laws have been challenged by other nations for inhibiting the free trade agreed to under the General Agreement on Tariffs and Trade.

5. Endangered Species Act of 1973

The 1973 Endangered Species Act (ESA) has a broad mandate to restrict the taking of species that are at risk of becoming extinct, protect and acquire the habitat necessary for those species to survive, regulate their trade, and force all federal agencies to evaluate the impacts of their activities on endangered species and to avoid jeopardizing a species' continued existence. This act was intended to enable the recovery of the populations of species it protects.

An endangered species is defined by the act as one that is in danger of extinction throughout all or a significant portion of its range. A threatened species is considered likely to become endangered in the foreseeable future throughout all or a significant portion of its range. The act outlines detailed procedures for the secretary of interior or commerce to follow in listing or delisting a species as threatened or endangered. The decision must be based on objective, scientific information on the biological status of the species and not on its economic or other value. Until a species is listed, it does not receive most of the protections of the act and the Department of Interior has a large backlog of species in the listing process.

The ESA outlines many duties and prohibitions that apply to all people subject to U.S. laws (in Section 9) and other duties that apply only to federal agencies (in Section 7). It is illegal to take any endangered or threatened animal species; take means to harm, harass, pursue, hunt, shoot, wound, kill, trap, or capture. Regulations further define harm to include habitat degradation to the point that wildlife is injured or killed. Although the original act did not prohibit taking of plants, later amendments prohibited taking plants on federal lands and prohibited removing or destroying plants on nonfederal lands where such action violated other laws.

Federal agencies are prohibited by Section 7 from approving, funding, or directly undertaking actions that are likely to jeopardize the continued existence of an endangered or threatened species or result in the destruction or adverse modification of its designated critical habitat. These agencies must consult with the Department of Interior or Commerce to determine the potential for jeopardy to the species or harm to its critical habitat and reasonable ways to avoid such harm. There is no direct legal obligation under ESA for nonfederal agencies or persons to avoid adverse modification of critical habitat. After the Supreme Court ruled that the Tellico Dam could not be completed due to the threat to the endangered snail darter (a fish), Congress amended ESA to allow a process for seeking exemption from the act's restriction on federal agencies.

Section 7 also requires designation of critical habitat for species listed as endangered or threatened. Critical habitat is defined as an area critical to the survival of the species; this could include a portion of the area a listed species occupies or an area beyond currently occupied habitat. Although the requirement to protect critical habitat has been applauded by many, it has been difficult to implement and designations of critical habitat by the Department of Interior or Commerce have been rare. The requirement has resulted in a great deal of litigation brought both by those challenging critical habitat designations and by those challenging the agencies for failing to make them.

The ESA can be enforced by the government through civil and criminal penalties and by citizens through citizen suits provided for in the act. Those who knowingly violate the ESA for an endangered species can be fined up to $50,000 and put in prison for up to 1 year. For a threatened species, a violation can result in a $25,000 fine and prison for 6 months. Additional civil penalties can be imposed. States can participate in enforcement of ESA by signing cooperative agreements with the secretary of interior and are also eligible for federal financial assistance for enforcement.

An international mandate is also included in the ESA. The act directs the secretary of interior to encourage other countries to protect endangered species and authorizes financial support and assistance from U.S. wildlife officers for such programs. Implementation of CITES is also authorized by ESA.

The ESA is one of the most controversial federal environmental laws passed during the past century. Its very specific, substantive, and legally binding provisions make it a target for continuous challenge from a variety of perspectives. Challenges have come from those who argue about whether ESA really helps protect species, inhibits use of private land, or applies to specific federal and private actions. The act forces attention to solving the problem of threatened and endangered species, with its wider attention to habitat conservation. This focus on solutions has forced discussion of how humans develop and use broad landscapes for economic production in areas with important biodiversity. More procedurally oriented laws are able to avoid controversy by not requiring that threats to biodiversity be addressed.

6. Fishery Conservation and Management Act (Magnuson Act) of 1976

Beginning in the early 1900s, international agreements between nations fishing in the same waters attempted to establish control of exploitation of ocean fish; by the 1970s the United States was a party to approximately 20 such agreements (Bean and Rowland, 1997). While waiting for agreement on a single international convention on fishing, the United States passed the Fishery Conservation and Management Act in 1976. The act established a 3-mile-wide territorial sea off the coast of the United States and a 200-mile-wide exclusive economic zone (EEZ). Within the EEZ the United States claimed exclusive management authority over fish and all other forms of marine animal and plant life except birds, mammals, and tuna. Tuna was added to this authority in 1990 amendments. The United States also claimed authority to manage certain sedentary species found in places where the continental shelf extends beyond 200 miles from shore.

The act required eight Regional Fishery Management Councils, which include state and federal officials, to develop comprehensive plans for conserving and managing the fisheries off their coasts. These plans have to comply with standards that include preventing overfishing and ensuring a maximum sustainable yield. The 1996 reauthorization of the act created new standards for minimizing "bycatch" (the accidental harvest of fish species not intended for harvest) and required that management plans try to minimize adverse impacts on fish habitats. Implementation authority for the act lies with the National Marine Fisheries Service in the Commerce Department. An important and controversial issue is the extent to which international trade agreements, such as the General Agreement on Tariffs and Trade or the North American Free Trade Agreement, have the potential to conflict with, and undermine, national laws aimed at protecting biodiversity.

C. Conservation of Habitats and Ecosystems

During the past 30 years, legislators, administrators and the general public have come to understand that species are not conserved without their habitats and ecosystems. The ESA reflected this by requiring conservation of critical habitat. Implementation of ESA has increasingly focused on habitat conservation plans as a way to conserve species. Many recent laws have focused on conserving lands and ecosystems in their entirety. These laws are not meant to replace species-oriented laws but rather to complement them. One of the challenges in dealing with ecosystem conservation is that ecosystems are naturally unstable and change continually; it can be difficult to separate natural background rates of change from changes that result from human disturbance.

One important and threatened habitat type, wetlands, received protection through the Clean Water Act of 1972. The dredge and fill permit program in Section 404 of the act is a significant force in protecting aquatic ecosystems by regulating the physical alteration of "waters of the U.S.," including wetlands, estuaries, and streams.

1. Federal Lands

Legislation and regulations addressing conservation of habitats and ecosystems in most countries focus primarily on creation of protected areas on publicly owned lands. On average, less than 5% of a country becomes legally designated as protected, and enforcement of that limited area of protection is often weak. Many protected areas allow some consumptive use of plant and animal species. An increasing number of countries have legislation that requires special permits to clear native vegetation or special habitat types outside of protected areas.

In the United States, approximately 30% of the land area (more than 700 million acres) is owned by the federal government and managed primarily by four agencies. Congress has enacted separate laws outlining how each agency should manage its lands. Historically, public land management agencies have given high priority to production of a commodity, such as timber, grazing land, or waterfowl production, and put less emphasis on maintaining the health of an ecosystem and its biodiversity. The 1964 Wilderness Act authorized Congress to designate portions of any federal land as wilderness areas, in which new commercial enterprises and permanent roads are prohibited and natural ecological processes are supposed to be allowed to occur without interference. During the early 1990s, federal agencies were directed to begin using "ecosystem management" to administer their lands. A level of production of a particular biological resource is to be considered in light of the health of the entire ecosystem.

Only the National Wildlife Refuge System (created in 1966) and the National Marine Sanctuary System (created in 1972) have as primary goals the conservation of wildlife and marine resources, respectively. The National Marine Sanctuary Act of 1972 authorizes the secretary of commerce to designate sanctuary areas to "maintain, restore, and enhance living resources" by providing critical habitat for endangered species and areas where ecosystem structure and function are maintained. The National Wildlife Refuge System Improvement Act of 1997 consolidated refuge-related authorities and stated the mission of the National Wildlife Refuge System. The system's mission is to administer a national network of land and water "for the conservation, management, and restoration of fish, wildlife, and plant resources and their habitats" and to maintain the system's biological integrity and diversity.

a. National Park Service Act (1916)

The National Park Service Act outlines the goals and administrative guidelines for conserving the scenery, natural and historic objects, and wildlife in the national park system (which includes parks, monuments, preserves, wild rivers, and lakeshores) in a manner that does not impair those resources for future generations. There have been many instances in which the National Park Service has had to balance the (occasionally) conflicting goals of wildlife management and human recreation. Amendments to the act in 1978 provided the secretary of interior with some authority to protect park resources from outside threats, such as logging on private lands adjacent to parks. Courts have also upheld National Park Service authority to prohibit or regulate fishing, trapping, and hunting within park lands.

b. National Forest Management Act of 1976

National forest lands began to be set aside in the late 1800s, with the purpose of protecting water supplies and timber production. The 1960 Multiple-Use Sustained-Yield Act indicated that the purpose of national forests is to provide for outdoor recreation, grazing, timber, watersheds, fish, and wildlife. The 1976 National Forest Management Act (NFMA) required the creation of management plans for each national forest. The plans are required to integrate physical, biological, and economic issues in order to provide for a diversity of plant and animal communities, including viable populations of native vertebrates. The plans are supposed to outline how timber harvests will mitigate impacts on biological diversity.

One of the most famous ecosystem management plans resulting from this legislation was for the old-growth forest in the Pacific Northwest. The Forest Service, after several court cases, was forced to develop a plan for old-growth forest that managed for a diversity of plant and animal species, including the spotted owl. Once the spotted owl was listed under the ESA, the management plan had to comply with both ESA and NMFA.

2. Private Land

Individuals and institutions privately own approximately 60% of the U.S. land area. Many of the previously discussed laws, such as the ESA and Section 404 of the Clean Water Act, aim to regulate actions with adverse impacts on biodiversity on private lands. Other federal laws aim to create positive financial incentives for conservation on private lands.

The 1985 Food Security Act included creation of a Conservation Reserve Program with the goal of encouraging farmers to voluntarily remove marginal lands from intensive agricultural production and plant them in perennial grass or tree cover for at least 10 years. A farmer enters into a contract to keep land in the reserve program for 10–15 years in exchange for cash or commodities. In addition, the farmer might receive up to 50% of the cost of certain conservation practices, such as creating wildlife corridors.

Biodiversity on private lands can also be conserved through donation or purchase of conservation easements. An easement is an interest in land that restricts the owner's use of the property in some way. The easement holder, usually a private interest or a government, has the right to enforce the restriction in perpetuity or for an agreed on time period. Since the 1970s, the Fish and Wildlife Service has been negotiating easements on wetlands as waterfowl production areas. The 1990 Food, Agriculture, Conservation and Trade Act created a Wetlands Reserve Program through which the government purchases easements on wetlands from farmers, who then restore and protect the functional values of the wetland. The farmer is eligible to receive a share of the costs of implementing the land management plan for the easement. A similar Environmental Easement Program was created at the same time. Under this program the government purchases easements on areas critical for wildlife on farms and ranches. The Forest Legacy Program, created in 1990, allows the secretary of agriculture to purchase conservation easements on forest areas that are threatened with development for other uses.

The 1969 U.S. Internal Revenue Code began permitting landowners to take a charitable tax deduction for qualified conservation easements, thus creating an incentive for private landowners to protect their lands. Protection of natural habitat for plants and animals is one conservation purpose that qualifies an easement for tax deduction. This legislation has led to a dramatic increase in donation of conservation easements on private land to nonprofit land conservation organizations.

The 1985 and 1990 Farm Bills contained a strong incentive, in the "swampbuster" provisions, for farmers to maintain wetlands and not convert them to agricultural production. Anyone who drained, dredged, filled, or altered a wetland in order to produce an agricultural product would lose their eligibility for federal price supports, crop insurance, the Conservation Reserve Program, and all other federal benefits for all of their land and agricultural products. The 1996 Farm Bill weakened this incentive, but swampbuster still serves to discourage conversion of wetlands to agricultural production.

D. Other National Laws

1. National Environmental Policy Act of 1969

The most significant provision of the National Environmental Policy Act (NEPA) for biodiversity is the requirement that federal agencies prepare an Environmental Impact Statement (EIS) for every federal action or major legislative proposal that can have a significant impact on the environment. The EIS is expected to address potential impacts of proposed programmatic or site-specific actions on the environment, including biodiversity. An important feature of NEPA is the requirement that there be a public comment period on the draft EIS, and that public concerns be addressed in the final EIS. Most federal agencies evaluate the effects of their programs and projects on threatened and endangered species and sensitive habitats. However, many EISs have given inadequate attention to impacts on less threatened species, to overall habitat impacts, and to cumulative impacts on biodiversity.

2. Foreign Assistance Act of 1962

The 1962 Foreign Assistance Act (FAA) authorized the creation of the U.S. Agency for International Development (USAID) and had little to say about environmental issues. Many of the areas of highest species diversity are located in developing tropical nations with insufficient economic resources to adequately protect biodiversity. As scientists and U.S. citizens became aware of the global loss of biodiversity in the late 1970s, they began to lobby the U.S. Congress to amend the FAA to ensure that U.S. foreign aid to other countries helped to conserve biological diversity and tropical forests. Beginning in 1982, amendments to the FAA have required that the United States assist developing nations with protection of habitats in protected areas, antipoaching measures for species, and research on biodiversity. Many of the most important and threatened national parks in developing countries now receive a significant portion of their operating budgets from bilateral assistance provided by developed nations.

Beginning in 1979, Section 216 of the FAA required that all U.S.-funded development projects conduct environmental impact assessments to ensure that tropical rain forests and other biologically important habitats were not adversely impacted. USAID is also required to monitor the impacts of other multilateral donor projects, particularly those funded by the World Bank, and to provide input to the U.S. representative to the World Bank about projects of environmental concern.

IV. STATE AND LOCAL LEGISLATION

The brevity of the following description of state and local laws does not imply that state and local legislation and regulations are unimportant to conservation of biodiversity. In fact, the opposite is true. The following brief sections merely skim the surface of the laws for 50 states and thousands of municipalities in the United States.

A. State Laws

All of the international and national laws and regulations outlined previously have their impacts at local levels. Primary responsibility for stewardship of biological resources is reserved to states by the 10th Amendment to the U.S. Constitution, which states that all powers not given to the federal government by the Constitution are retained by the states. Federal laws, such as those described previously, define minimum standards for the states to follow. States can, and sometimes do, pass laws that are more restrictive than federal laws in protecting plant and animal species, wetlands, forests, and other ecosystems.

States own their wildlife resources and generally attempt to manage them to benefit all citizens. Historically, most states have focused their wildlife conservation activities on a relatively small number of hunted or fished species. States retain authority to regulate and legislate many issues related directly to conservation of biodiversity; for example, managing the natural resources on state-owned lands, setting hunting limits on deer and other resident, nonmigratory species, and approving permits for development and other projects that affect many habitats and species. Perhaps the largest impact of state laws and local ordinances on biodiversity relates to how well the laws regulate the approval of economic development plans so as to protect natural habitats and mitigate the impacts of development. Many states also require that municipalities develop comprehensive plans, which might address conservation of biological resources along with land use and development issues.

Only a small number of states have passed laws and regulations specifically addressing conservation of genes, species, and ecosystems, although the number is increasing. In 1992, the state of Michigan passed a Biological Diversity Conservation Act that required creation of a state biodiversity conservation strategy. A 1993 statute in New York State established a Biodiversity Research Institute with a mandate to coordinate state and private efforts to collect information about the state's biodiversity. The law also directed the New York Department of Environmental Conservation to implement programs to conserve rare and endangered species on state lands. In 1996, the state of Wisconsin passed a bill that revised the state's forest statute to ensure that management of state forests is consistent with maintaining native biodiversity and sustainable ecosystems. Many states have statutes that authorize creation of conservation easements on private lands. Hawaii created a statutory tax incentive that provides owners of forest or water reserve lands with relief from all property taxes in exchange for allowing the state to manage their lands (Defenders of Wildlife, 1996). Hawaii also has strong statutes to control the introduction of exotic species.

Many states, including California and Tennessee, have created executive orders and memoranda of understanding regarding conserving biodiversity, but these do not have the force of law. Many other states have policies aimed at conserving biodiversity by creating land acquisition programs for areas with important biodiversity, prohibiting introduction of exotic plant and animal species, and requiring environmental impact assessments in order to minimize and mitigate the impacts of development projects on biodiversity.

Many states are involved in legislatively mandated, regional initiatives to conserve ecosystems that cross state boundaries. The 1987 Chesapeake Bay Agreement was signed by Virginia, Pennsylvania, Delaware, Maryland, the District of Columbia, the Chesapeake Bay Commission, and the U.S. Environmental Protection Agency. The signatories agreed to work on a program to conserve the biological resources and water quality of the Chesapeake Bay system.

B. Local Ordinances

Local laws can be enacted by municipalities, towns, cities, counties, townships, or boroughs, depending on their regulatory authority and on what local government units are called in different parts of the United States. Although ultimately all biodiversity is conserved or lost at a local level, rarely are local laws primarily concerned with biodiversity per se; rather, they tend to focus on biological resources that indirectly influence public health, safety, or welfare (e.g., protecting wetlands or wooded slopes to control floods or regulating hunting for safety or to maintain economic benefits). Increasingly, however, concerned citizens have been working to ensure that local laws do consider conservation at some scale. Local governments can address biodiversity through broad municipal comprehensive planning and through municipal ordinances that address specific human activities that affect biodiversity on a more site-specific level.

Whether required by state law or not, many municipalities create comprehensive plans that address future land uses in the context of economic growth and quality of life, including conservation of biological resources. Many comprehensive plans include sections on natural or environmental resources. Some municipalities are beginning to prepare separate plans for "open space"—undeveloped land that often contains natural habitats. Most open space plans identify specific environmentally important habitats and recommend specific conservation strategies for those habitats.

Municipalities have adopted a variety of ordinances intended to minimize impacts on species and natural habitats; often, these ordinances support the resource conservation goals established through comprehensive and/or open space plans. An increasing number of municipalities are enacting ordinances that offer stronger protection to stream and wetland ecosystems than the requirements of state and federal law. Some municipal ordinances focus on integrating protection and restoration of habitats with land development project approvals. For example, some local ordinances have required that development projects maintain or create natural vegetated buffers of more than 25 ft along wetland boundaries and more than 75 ft along each side of a stream.

An increasing number of municipalities have passed forest conservation ordinances that limit development in woodlands and penalize developers who harm those habitats. Some municipalities will not approve logging permits for sensitive woodlands without implementation of a forest management plan that demonstrates sustainable practices such as leaving buffers along streams, restricting logging to seasons when birds are not nesting, or leaving dead trees standing for wildlife habitat. Forest conservation ordinances have required developers to demonstrate that their plans will minimize unnecessary forest fragmentation and will not exceed the percentage of any woodland allowed to be cut. Plans that exceed the legal cutting percentage might only be approved if the developer mitigates the impact through reforestation or off-site forest conservation. Some ordinances seek to minimize tree removal by setting a limit on clearing around buildings and roads (e.g., 25 ft).

Many local municipalities in agricultural areas have ordinances that prohibit growing of introduced "weeds" that might decrease the productivity of agricultural crops. In some residential municipalities, weed laws have been replaced with natural landscaping ordinances that promote and govern natural alternatives to lawn grasses, such as wildflower meadows and reforestation.

V. CONCLUSIONS

The evolution of government legislation and regulation of biodiversity during the past 100 years reflects both the dramatic increases in human impacts on biodiversity in every part of the earth and the increased understanding of the structure and function of the natural world. Legislation has evolved from individual nations focusing on single species of economic value to multinational agreements aimed at protecting entire habitats and ecosystems. The three greatest current threats to biodiversity are habitat alteration and destruction, overexploitation of species, and introduction of exotic species; the laws summarized previously address these threats to varying degrees. Despite continuous significant improvements in legislation and regulations aimed at conserving biodiversity in the face of economic development, genes, species, and habitats continue to be lost at an increasing rate.

There are many challenges to the effectiveness of the laws and regulations described in this article. Perhaps the greatest challenge is ensuring that existing laws are implemented as intended and then enforced. Many laws require scientific information on the status of species and ecosystems; such information is often unavailable or extremely costly to obtain. In the United States, the federal agencies implementing laws such as the ESA are frequently so underfunded by Congress that they cannot achieve their conservation goals. It is also a challenge to ensure that state and local agencies receive the funding required to implement conservation laws. Federal, state, and local implementing agencies are often pressured by special-interest groups to compromise conservation mandates in favor of short-term economic gains. The laws in many countries go unenforced due to lack of oversight from private citizens and advocacy groups. Finally, most laws are based on prohibiting actions that harm biodiversity; there is still a need for more legislation that creates positive incentives for conserving and enhancing biodiversity. The need for positive incentives is particularly acute for private landowners, who control 60% of the land in the United States.

Although the laws and regulations regarding conservation of biodiversity have generally improved significantly during the past 30 years, it is important to acknowledge that they can be amended and weakened whenever the various levels of government, from international to local, choose to do so. In addition, many laws and regulations in other sectors create incentives to destroy biodiversity, such as logging subsidies, underpriced grazing rights, subsidies for economic development, and international trade policies.

Acknowledgments

Thanks to Jan Goldman-Carter, M. Lynne Corn, Michael Hill, and David Harper for reviewing drafts of this mss prior to publication.

See Also the Following Articles

CONSERVATION EFFORTS, CONTEMPORARY • CONSERVATION MOVEMENT, HISTORICAL • ENDANGERED ECOSYSTEMS • PROPERTY RIGHTS AND BIODIVERSITY

Bibliography

Bean, M. J., and Rowland, M. J. (1997). *The Evolution of National Wildlife Law*, 3rd ed. Praeger, Westport, CT.

Council on Environmental Quality (1993). *Incorporating Biodiversity Considerations into Environmental Impact Analysis under the National Environmental Policy Act.* Council on Environmental Quality, Washington, D.C.

Defenders of Wildlife (1996). *Saving Biodiversity: A Status Report on State Laws, Policies and Programs.* Defenders of Wildlife, Washington, D.C.

Environmental Law Institute (1998). Protecting biodiversity: Legal mechanisms concerning access to and compensation for the use of genetic resources in the United States of America. Research report. Environmental Law Institute, Washington, D.C.

Heywood, V. H., and Watson, R. T. (Eds.) (1995). *Global Biodiversity Assessment.* Cambridge Univ. Press, Cambridge, UK.

Houck, O. (1993). The Endangered Species Act and its implementation by the U.S. Departments of the Interior and Commerce. *Univ. Colorado Law Rev.* 64(2), 277–370.

Kubasek, N. K., and Silverman, G. S. (1994). *Environmental Law.* Prentice Hall, Englewood Cliffs, NJ.

GRASSHOPPERS AND THEIR RELATIVES

Piotr Naskrecki
University of Connecticut

GLOSSARY

brachypterous Having tegmina and wings shorter than the abdomen but overlapping or touching each other on the dorsum.

cerci Paired, usually not segmented structures at the end of the abdomen, sometimes used by males during mating to grasp the female's abdomen.

hemimetabolous Having incomplete metamorphosis, that is, showing gradual change from molt to molt, with externally developing wing pads, and lacking any larval and pupal stages.

hypognathous Position of the head when the mouth is directed toward the ventral side of the body.

kHz A unit of frequency equal to 1000 hertz; the number of hertz (abbreviated Hz) equals the number of cycles per second (e.g., sound waves).

macropterous Having wings that are fully developed, reaching or exceeding the end of the abdomen.

micropterous Having wings that are greatly shortened, not overlapping or touching on the dorsum.

prognathous Position of the head when the mouth opening is directed forward.

spermatophore A membranous package containing sperm that is transferred from the male to the female during copulation.

stridulatory apparatus An organ of sound production based on the mechanism of rubbing one part of the body against another.

tegmina (singular: tegmen) Thickened forewings.

ORTHOPTERA IS THE ORDER OF INSECTS that includes grasshoppers, locusts, crickets, katydids, and their relatives. The name of the order is derived from the Latin word *orthos* (straight or rigid) and the Greek word *pteron* (wing), referring to the simple venation and position of their wings, which are always held straight and parallel to the main axis of the body. Most orthopterans are large or medium-sized insects. Body lengths of less than 10 mm are uncommon, while many exceed 50 mm in length, with some having bodies over 100 mm long and a wingspan of 200 mm or more.

I. OVERVIEW OF ORTHOPTERA

A. Introduction

Orthoptera are hemimetabolous insects, with nymphs resembling adult forms in their general appearance but lacking fully developed wings and reproductive organs. Mouthparts of orthopterans are of the chewing/biting type (= mandibulate). The head is hypognathous, rarely prognathous; the antennae are usually long and threadlike, consisting of fewer than ten to several hundred segments. The part of the body immediately behind the head, or the pronotum, is usually large, often shieldlike, and in extreme cases covers a large part (many katydids) or the entire body of the insect (pygmy grasshoppers).

The front and middle legs are cursorial (i.e. adapted for walking), yet in some cases the front pair of legs may be modified for digging (mole crickets, pygmy mole crickets, false mole crickets), or both the front and middle pairs may be modified for grasping (predatory katydids). In some orthopterans (most katydids and crickets) the front legs have tibial auditory organs (the ear). The hind legs of most orthopterans are saltatorial, (i.e. modified for leaping), with large, muscular femora and long, slender tibiae. Some grasshoppers can perform repeated leaps of 2.6 m without any obvious signs of fatigue. This is possible primarily because of the presence in their back legs a protein called resilin. Resilin has superb elastic properties, with 97% efficiency in returning stored energy. This allows for explosive release of energy that catapults the insect, a task impossible with muscle power alone. Certain groups of orthopterans, especially those leading a subterranean life, have lost their ability to jump and their hind legs resemble typical cursorial legs. In some grasshoppers and certain Ensifera the inner surface of the hind femur is modified for sound production (stridulation).

The wings of orthopterans are either fully developed or reduced to various degrees. Wing polymorphism, or the occurrence of individuals with well-developed and reduced wings within the same species, is not uncommon. The forewings are somewhat thickened, forming leathery tegmina. In most katydids and crickets parts of the tegmina are modified for stridulation. The hindwings, when present, are fanlike, hidden under the first pair in the resting position. Often the hindwings are longer than the tegmina and protrude behind their apices. Wing buds of nymphal stages are always positioned in such a way that the second pair of wings overlaps the first one, whereas in adult individuals of micro- and brachypterous species, the first pair of wings always overlaps the second pair, despite their nymphal appearance. The base of the abdomen in grasshoppers has lateral auditory organs known as abdominal tympana. Females of most orthopterans have a prominent ovipositor at the end of the abdomen, derived from the eight and ninth abdominal segments. Katydids and crickets usually have a well-developed ovipositor—sword-, sickle-, or needle-shaped—whereas females of grasshoppers and their relatives usually lack a long, external ovipositor.

The number of described species of Orthoptera is approximately 25,000. This number, however, probably represents only a half or less of the actual number of species of Orthoptera present on earth today. Tropical regions of South America, Africa, and Asia still remain virtually unexplored in terms of their grasshopper and katydid fauna, and many thousands of species are expected to be described in the future. The Australian orthopteran fauna is the best studied of all tropical regions of the world, yet more than 1500 Australian species still remain to be formally described despite being already recognized as new to science.

Members of the order Orthoptera inhabit virtually all terrestrial habitats of the world, from rock crevices of the littoral zone of the oceans, subterranean burrows, and caves, to treetops and peaks of the alpine zones of mountain ranges. Both deserts and grasslands as well as dense forests have rich and unique orthopteran faunas. There are few purely aquatic forms, but many are associated with marshes and other semiaquatic habitats. Orthopterans are important members of nearly all terrestrial ecosystems, both in the role of consumers and prey. Massive outbreaks of some species of grasshoppers (and less frequently katydids and crickets) can cause enormous losses for the food industry and forestry. Locusts (a vernacular name for certain grasshopper species that tend to produce large seasonal outbreaks, and not a taxonomic entity) have been a part of human history from the very beginning of our agricultural tradition. They still pose a great risk for agriculture in many parts of the world, although they pose less of a problem now than a few hundreds years ago, thanks mostly to better understanding of their population dynamics and application of various chemical and biological control measures.

B. Sound Production

The single characteristic most frequently associated with grasshoppers and their relatives is their ability to produce sounds. Although less widespread than generally believed, this ability is nonetheless quite common in some groups of orthopterans. The role of sound production is three-fold and similar in some respects to that of birdcalls: (a) attraction of mates, (b) territoriality, and (c) disbursement of release calls (alarm calls produced when seized by a predator). The calls of orthopterans are usually species-specific and play a very important role in species recognition. The information in the call may be coded in the form of frequency modulation (the pitch of the call changes through time, a mechanism best known in birds), time modulation (the pitch of the call remains the same throughout its duration but its temporal pattern is unique to the species), or both modes combined.

The dominant mechanism of sound production in Orthoptera is stridulation, which involves rubbing one modified area of the body against another. Contrary to popular belief, no orthopterans (nor any insect for that matter) produce sound by rubbing their hind legs against each other. Katydids (Tettigoniidae) and crickets (Grylloidea) produce sound by rubbing a modified vein (the stridulatory vein) of one tegmen against a hardened edge of the second tegmen (the scraper). The stridulatory vein is equipped with a filelike row of teeth, the number of which varies from a few to a few hundred. In most katydids, the stridulatory area is situated at the base of the tegmina, except in brachypterous species where it covers their entire surface. In crickets virtually the entire surface of the tegmina is modified for stridulation. As a rule, in katydids the stridulatory file is situated on the left tegmen and the scraper on the right one, while in crickets the situation is reversed. A membranous area at the base of the tegmen, the mirror, amplifies the sound. In addition, some katydids (e.g. *Thoracistus*) use their enlarged, shieldlike pronotum as an additional sound amplifier. Crickets, lacking the enlarged pronotum, use other methods of sound amplification, such as singing from burrows, the shape and size of which is attuned to boost certain frequencies (*Gryllotalpa*), or using the surface of a leaf for the same purpose (*Oecanthus*). The ability to stridulate is restricted almost exclusively to males, although in a few groups of katydids (Phaneropterinae, Ephippigerinae), females respond to the male's calls by stridulating as well. Their sound apparatus is not homologous to that of the males and is usually quite simple, lacking the sophisticated mechanism for sound amplification. In addition to tegminal sound apparatus, a few groups of katydids have developed other mechanisms of stridulation. For example, all members of the Australasian subfamily Phyllophorinae lack the typical wing stridulation and produce sound, instead, by rubbing their hind coxae against modified thoracic sterna. Mandibular sound production occurs in some members of Mecopodinae.

Grasshoppers use the same principle of stridulation, but instead of rubbing their tegmina against each other, these insects produce sound by rubbing the inner surface of the hind femur against one of the veins of the tegmen. In the slant-faced grasshoppers (Gomphocerinae), the inner surface of the femur possesses a file of small knobs while the vein on the tegmen acts as the scraper. In band-winged grasshoppers (Oedipodinae), the vein has a row of pegs and the femur plays the role of the scraper. In addition to these two principal mechanisms, some grasshoppers stridulate by rubbing their hind legs against the sides of the abdomen (Pamphagidae) or by kicking their legs feet against a modified area at the apex of the tegmen (*Stethophyma*). Australian false mole crickets (Cylindrachetidae) have a stridulatory file at the base of their maxillary palps, and some species of pygmy mole crickets (Tridactylidae) produce sound by rubbing a modified vein on the dorsal side of the tegmen against another vein at the base of the hindwing.

The sound frequencies produced by orthopterans during stridulation vary from a few kHz (most crickets and grasshoppers) to well above 100 kHz (some katydids). Crickets' calls are characterized by their tonal purity, with most energy of the call allocated within a narrow range of frequencies. Katydid calls vary from tonally pure (although often well above the human hearing range) to broad, noiselike signals. Grasshoppers produce mostly broad spectrum, noiselike calls. Unlike many vertebrate calls, many orthopterans produce time modulated rather than frequency modulated signals. Crickets are a notable exception and most species produce melodious, birdlike, frequency-modulated chirps.

In addition to stridulation, some grasshoppers crepitate in flight. In this case the sound is produced by hitting certain veins of the wings against other veins. This behavior is especially common among band-winged grasshoppers (Oedipodinae) and plays an important role in courtship and territorial displays.

A few members of normally acoustic orthopterans have lost their ability to produce airborne signals and instead have developed a number of substitute mechanisms of substrate-borne communication. Males of the oak katydids (*Meconema*) lack the typical tegminal sound apparatus and instead produce sound by drumming with their hind legs against the bark of trees. Similar drumming behavior, although still accompanied by typical stridulation, is a component of the courtship behavior of the pitbull katydid (*Lirometopum*). Males of the cricket *Phaeophilacris spectrum* have lost their ability to stridulate and instead signal by rapidly flicking their tegmina back and forth while holding them in vertical position. The near-field motion is detected by female's cerci, rather than her ears. Despite having a fully developed stridulatory apparatus, many neotropical members of the katydid subfamilies Pseudophyllinae and Conocephalinae spend little or no time stridulating, relying instead on substrate-borne tremulations. In this case, a male stands rigidly on a leaf or stem of a plant and violently shakes his entire body. The low frequency waves are transmitted along the branches of the plant. The reason for this behavior is unclear, although a few hypotheses trying to explain it have been proposed. The most widely accepted one is the avoidance of predation by foliage-gleaning bats that are known to use insect sounds to locate their prey. Others include eluding satellite males (nonsinging males of the same species trying to intercept a female), avoidance of parasitoid flies, and helping females locate males on multibranched plants. Thanks to the rapid development of recording techniques in recent years, it is now known that many groups of orthopterans previously believed to be silent appear to employ a number of techniques of substrate communication.

C. Reproduction and Growth

The courtship and mating behavior of orthopterans provide some of the most complex and fascinating spectacles of the insect world. In addition to sound production, as discussed earlier, many species employ visual, tactile, and olfactory signals in their mating strategies. Visual communication is especially well developed in grasshoppers, where males often have bright, species-specific markings on different parts of their bodies, displayed in carefully choreographed sequences during courtship. Grasshoppers of the genus *Syrbula* are definite champions in this respect, and males of some species, in addition to calling, perform a dance consisting of 18 distinct movements. Visual signals employed by many diurnal grasshoppers include flight displays, where males flash their colorful hindwings (this is sometimes accompanied by crepitation), flagging with distinctly colored hind legs, and displays involving brightly colored, and often enlarged antennae. Courtship of katydids and crickets relies less on visual signals and more on sound and chemical cues, which are more appropriate for these mostly nocturnal animals. In both groups, males sometimes produce two different types of calls, a long-range advertisement call and a quieter, courtship song, performed only in the presence of a female. Female in some species may reply using either airborne signals or tremulation.

Chemical communication in Orthoptera is little studied, but there is evidence that at least some species employ it during courtship. Females of the New Zealand giant weta *Deinacrida rugosa* produce a musky substance that males use to locate females, while males of camel crickets of the genus *Ceuthophilus* posses thoracic glands that may also play a role in courtship. Field crickets *Teleogryllus comodus* use pheromones covering the female antennae to initiate courtship. Some other crickets use airborne pheromones in locating members of the opposite sex.

Copulation in orthopterans involves transfer of a sperm sac (spermatophore), which in some groups is accompanied by a large packet of nutritious proteins, the spermatophylax. Males of some orthopterans also allow females to feed on parts of their own bodies during copulation. Males of haglids (*Cyphoderris*) have their hindwings modified into thick, fleshy lobes, the sole purpose of which is to be eaten by the female during copulation. Females of tree crickets (*Oecanthus*) feed on males' thoracic glands during copulation, and in some crickets of the subfamily Nemobiinae the females feed on enlarged spines on males' hind tibia. Males of other orthopterans, lacking such tasty incentives, must rely on their strong grasp or modified cerci at the end of the abdomen to hold the female during copulation.

Oviposition takes place in a variety of substrates, such as soil, plant tissues, or rock crevices. In some cases, eggs are protected from desiccation by a foamy mass produced by the female. Nymphs usually hatch within a few weeks or months, but sometimes the eggs undergo a year long, or longer, diapause. Few orthopterans display any kind of parental behavior, although some crickets (*Anurogryllus*) lay eggs in burrows guarded by the female. Female mole crickets (*Gryllotalpa*) not only lay eggs in special egg chambers underground, but also actively care for the eggs by licking and removing fungal spores from their surfaces. The hatchlings stay with their mother for a few weeks before dispersing.

D. Food and Feeding

Orthopterans are extremely diverse in their food preferences and feeding techniques. Virtually all Caelifera (grasshoppers and their allies) are strictly herbivorous, very rarely engaging in cannibalistic behavior, and doing so only under crowded conditions. Most grasshopper species seem to be polyphagous (feeding on a wide variety of plant species), but some are oligophagous (feeding on a narrow spectrum of plant species) or monophagous (feeding on only one species of plants). The last are often associated with toxic, alkaloid-rich plant species, and these substances make the insect themselves inedible to many potential predators. Pygmy grasshoppers (Tetrigidae) are some of the few insects that feed on mosses and lichens.

Ensifera (katydids, crickets, and their relatives) range from herbivorous to omnivorous to strictly predaceous. Some katydids specialize on rather unusual food sources. The members of the Australian genus Zaprochilus feed exclusively on the pollen and nectar of flowers. Most katydids, however, feed on a wide range of organic material. For example, the Central American Rhinoceros katydid (*Copiphora rhinoceros*) is known to feed on flowers, fruits, hard seeds, caterpillars, other katydids, snails, frog eggs, and even small lizards. Strictly predaceous katydids employ both the "sit-and-wait" strategy (Saginae) or actively forage and hunt living insects (Listroscelidinae). Some raspy crickets (Gryllacrididae) also actively search for insect prey by rapidly running along branches and grasping any sitting insect their encounter. Crickets and cave crickets tend to be generalists in their dietary preferences but rarely exhibit tendencies to feed on live prey. Some mole crickets have a behavior unique among orthopterans, and insects in general, of gathering and storing germinating seeds in circular chambers below ground for later consumption.

E. Phylogeny and Taxonomy

Despite a long history of taxonomic research on orthopteroid insects and despite their economic importance, there is little agreement as to the taxonomic position of many groups of the Orthoptera. Unfortunately, the current classification system is still based mostly on 19^{th}-century works. Older classifications also included in this order such groups as cockroaches (Blattodea), praying mantises (Mantodea), walking sticks (Phasmodea), and earwigs (Dermaptera), each of which is now considered a separate, albeit closely related order. For the purpose of this review, the classification system proposed by Rentz (1996) will be adopted. This system divides the order Orthoptera into two suborders: Ensifera, with superfamilies Tettigonioidea (katydids), Grylloidea (crickets), and Gryllacridoidea (cave crickets and relatives); and Caelifera, with superfamilies Eumastacoidea (monkey grasshoppers), Acridoidea (grasshoppers and locusts), Tetrigoidea (pygmy grasshoppers), and Tridactyloidea (pygmy mole crickets and relatives). The following differences generally allow Ensifera and Caelifera to be distinguished (note that there are numerous exceptions):

Ensifera (katydids, crickets and relatives)	Caelifera (grasshoppers and relatives)
Antennae long and thin, with over 30 segments	Antennae short, with less than 30 segments
Anterior tibiae usually with tympanal organs	Anterior tibiae without tympanal organs
Ovipositor long, sword- or needlelike	Ovipositor very short
Sound produced by tegminal stridulation	Sound produced by rubbing hind legs against tegmina

The monumental task of describing all living species of Orthoptera is far from complete. As the exploration of tropical regions of the world progresses, more and more species awaiting to be formally described accumulate in museum collections worldwide. Unfortunately, at the same time the number of specialists capable of such work is dwindling rapidly.

F. Paleontology

The Orthoptera is an old group, dating back to the Carboniferous period. Along with Titanoptera and Phasmodea (= Phasmatoptera), they represent the superorder Orthopterodea. Both Titanoptera and Phasmodea branched off of the main stock of Orthopterodea in the early Permian. Titanoptera disappeared at the end of the Triassic while Phasmodea (walking sticks and their relatives) and Orthoptera still flourish today. The two main lineages of Orthoptera, Ensifera and Caelifera, probably separated in the late Carboniferous. Most modern families of Ensifera appeared between the early Jurassic and the early Triassic periods. The oldest, still extant family of Ensifera, the Prophalangopsidae, appeared in the early Jurassic. The oldest family of Caelifera, the Eumastacidae, appeared in the middle Jurassic, followed by the Tetrigidae and the Tridactylidae at the beginning of the Cretaceous.

II. TETTIGONIOIDEA (HAGLIDS AND KATYDIDS)

A. Introduction

The superfamily Tettigonioidea includes two families: the Prophalangopsidae (haglids or hump-winged crickets) and the Tettigoniidae (katydids or bush-crickets) (Fig. 1). It is the second largest group of Orthoptera (grasshoppers being more speciose), with nearly 6300 described species assigned to nearly 1100 genera and 20 subfamilies. Tettigonioidea are characterized by the presence in the males of well-developed tegminal stridulatory apparatus; long, threadlike antennae, usually longer than the body; four-segmented tarsi; short, nonarticulated cerci in the males; and a large external ovipositor in the females. The wings, when present, are usually held in a roof-like fashion over the abdomen. Virtually all members of the superfamily have bilateral tympanal organs on the front tibia as well as an additional acoustic trachea on the thorax behind the pronotum.

The Tettigonioidea are distributed throughout the world, but like most other orthopterans, they show their greatest diversity in the tropical and subtropical regions. Although many groups of katydids are quite diverse in temperate regions of the world, they are less common in areas with long, cold winters. A large proportion of the Tettigonioidea are arboreal but many also inhabit grasslands, savannas, deserts or rocky mountain tops. Most species are nocturnal or crepuscular, though several major groups consist almost exclusively of diurnal species.

FIGURE 1 Tettigonioidea. *Pycnopalpa bicordata* from Costa Rica (Tettigoniidae, Phaneropterinae). See also color insert, Volume 1.

B. Major Lineages

1. Prophalangopsidae (Haglids or Hump-Winged Crickets)

This ancient family, dating back to the early Jurassic period, is represented by only six extant species. They are characterized by a rather primitive stridulatory apparatus lacking a proper mirror, head with antennae inserted near the lower margin of the eyes, and the metatarsus with a well-developed pulvillus. The rarest of these living fossils, *Prophalangopsis obscura*, is known from a single specimen collected over a hundred years ago at an unknown location in India, and is the sole member of the subfamily Prophalangopsinae. Naturally, little is known about this insect other then its strong resemblance to forms extinct millions of years ago. The remaining five species belong to two genera of the subfamily Cyphoderrinae and are distributed in western North America (*Cyphoderris*) and northeastern Asia (*Paracyphoderris*). The biology of *Cyphoderris* is quite well known and some species of the genus are often used in studies on mating behavior and parental investment in offspring. These insects are unique among the Tettigonioidea in the fact that the male's stridulatory apparatus is almost perfectly symmetrical—that is, males have fully functional files and scrapers on both tegmina and can produce sound with either the left or right tegmen (hence the name "ambidextrous crickets," sometimes applied to these insects). The hindwings are reduced to fleshy lobes, which are devoured by the female during copulation. Males who had already mated once and are missing these courtship "snacks" must resort to other methods of holding the female's attention and instead use the "gin trap," a complex system of cuticular modifications whose role is to hold the female's abdomen firmly in place during copulation. These insects are also unique in their ability of being active, and even call during winter, from snow-covered bushes, when the temperature oscillates around 0°C.

2. Tettigoniidae (Katydids or Bush-Crickets)

Katydids are a large and diverse group of orthopterans, elevated by some authors to the rank of a superfamily with numerous families. Nearly 6300 species have been described, placed in nearly 1100 genera and 18 subfamilies. Most katydids are medium sized or large insects. The largest katydid, *Siliquofera grandis*, from New Guinea, has a wingspan of over 250 mm. On the other hand, members of the Australian subfamily Microtettigoniinae are less than 10 mm long.

Males of almost all species have a well-developed stridulatory apparatus at the base of the tegmina. The exceptions are the subfamilies Phyllophorinae, Phasmodinae, and some members of Meconematinae and Mecopodinae. Males of the Australasian subfamily Phyllophorinae, despite having fully developed wings, lack the stridulatory apparatus and instead produce sound by rubbing their hind coxa against specially modified thoracic sterna. Australian Phasmodinae lack wings altogether and are presumably totally silent. This is also the only subfamily of katydids having greatly reduced tympanal tibial structures ("the ear") and lacking the thoracic auditory spiracle. All katydids have long, thin antennae, often longer than the body, in extreme cases exceeding its length three to four times. The head of katydids is hypognathus, rarely prognathous (Phasmodinae, Zaprochilinae), and sometimes bears distinct, cuticular, hornlike processes between the antennae. Their function is uncertain but is likely related to the fact that many predators (especially bats and birds) tend to seize their katydid victims by the head. The pronotum is enlarged, saddle shaped, and in some forms greatly expanded, covering part of the wings and the abdomen. In the male, the enlarged pronotum often acts as a call amplifier. Curiously, the group with a particularly enormous, boxlike pronotum, the Phyllophorinae, lacks the ability to produce sounds with their wings. The pronotum is also sometimes armed with sharp spines and processes. The wings of katydids can be fully developed or reduced to a varying extent, in some lineages (Tettigoniinae) showing significant intraspecific polymorphism in the degree of their development. The tegmina are sometimes reduced to scalelike organs, the only function of which is sound production. Hindwings in such taxa are either totally missing or greatly reduced. Many katydids use their wings to enhance their remarkable mimicry of leaves, twigs, or bark to a degree that can fool professional biologists. Members of the neotropical tribe Pterochrozini are particularly excellent mimics of both live and dead leaves, perfectly imitating the look and feel of leaves, including simulated traces of herbivory, with simulated lichens and mosses "growing" on them. A few katydids are Batesian mimics of wasps.

The legs of katydids are usually long and slender. Predaceous species frequently have rows of long spines along the ventral edges of the front and middle legs, using them for grasping and holding prey. In some Listroscelidinae the spines are particularly long, forming a kind of a net, used for scooping small flies and other insect prey. The hind legs are saltatorial, except for the members of some groups living in burrows or under rocks (Hetrodinae, Bradyporinae). The abdomen of males has a pair of cerci, in some groups developed to grasp the end of female's abdomen during mating. The ovipositor of females is usually long and sickle or sable shaped. The reduction of the ovipositor to short processes appears independently in ground-dwelling forms (Hetrodinae) and some arboreal ones (Phaneropterinae), apparently related to the fact that the eggs are laid on the surface of the soil or leaves, rather than inserted into them, as most katydids do.

The significant male investment in offspring is one of the characteristics of katydids that attracts many researchers to study their behavior and evolution. During copulation, males of many katydids produce a large, protein-rich spermatophylax, which is eaten by the female after copulation. The size of the spermatophylax can approach 60% of the male's body mass, making it an extremely costly and significant contribution to egg production. This causes the males of many species to be quite choosy in selecting their mating partners, and under certain circumstances, the females may compete for males, a role reversal remarkably rare in the animal world.

The geographic distribution of katydids reveals several large centers of endemism, the largest being Australia, where 5 out of the 13 subfamilies known from the continent are unique to it. The neotropics have the highest number of described species (approximately 1800) but no endemic subfamilies, followed by Eurasia (approximately 1350 species), the Indo-Malaysian region (approximately 1150), and Africa (approximately 850 species, 2 endemic subfamilies). These numbers, however, are very likely to be multiplied in the future if taxonomic work in these regions is intensified.

a. Major Subfamilies

Bradyporinae. Most likely the most basal subfamily of katydids, it is characterized by having a large, stout body, greatly reduced wings, and the tarsus with a metatarsal pulvillus. Members of this family are restricted in their distribution to southwestern Palaearctic. Often grouped with several other subfamilies (Hetrodinae, Ephippigerinae), they nonetheless appear to be a separate clade of their own. Only about 50 species have been described. All species of the subfamily are dark colored, often black, and resemble giant crickets rather than typical, graceful katydids. Most live on the ground or low vegetation. Some species are known to produce defensive autohemorrhage and squirt their hemolymph from orfices on their body if the insect is seized by a predator.

Phaneropterinae. This is the largest subfamily of Tettigoniidae, with more than 2100 species described from all continents of the world (other than the polar regions). Members of this subfamily are medium size to large, usually green and leaflike katydids, characterized by a lack of lateral grooves on the tarsi, and rather primitive venation of the wings. Wing reduction is widespread in several lineages of the subfamily but the stridulatory apparatus is always present. Female stridulation is known in a few genera. Virtually all species are phytophagous, although opportunistic cannibalism has been observed in a few species. Many species are exclusively arboreal. Eggs are laid in the soil, plant tissues, or on the surfaces of leaves and bark. Immature stages of many species mimic ants, tiger beetles, and even spiders, while adults of the same species usually mimic leaves or blades of grass. Several South American genera are superb mimics of pompilid wasps.

Acridoxeninae. This is a small, aberrant subfamily of large Central African katydids, with only one species, *Acridoxena hewaniana*. Its relationship to other katydids is uncertain as it displays a mixture of very advanced and very primitive characters. The most interesting feature of *Acridoxena* is its astonishing mimicry of a dried, spiny plant. All parts of the body look like shriveled, twisted, dry leaves or twigs. Nothing is known of its biology or behavior.

Ephippigerinae. This is a small Palaearctic subfamily, with about 130 described species. All species have greatly reduced, scalelike wings, usually hidden under an enlarged, saddle-shaped pronotum. Females of many species stridulate and certain species have long been model organisms for studies on acoustic communication and courtship in insects. Some species occasionally cause minor agricultural damage.

Pseudophyllinae. This is the second largest subfamily of Tettigoniidae, with more than 1100 described species, distributed in tropical and subtropical regions of the globe. A notable exception is the true katydid (*Pterophylla camellifolia*) of the eastern United States, which is both the northernmost member of the subfamily and the source of the name "katydid." Its characteristic, loud call resembles (although some listeners disagree) the syllables "ka-ty-did." Most pseudophyline katydids are found in the tropical areas of South America and Southeast Asia. Many are spectacular mimics of leaves (tribes Pterochrozini and Pseudophyllini) and bark (tribes Plaminini and Cymatomerini). Secondary loss of stridulation and widespread presence of tremulation is characteristic of most neotropical members of the subfamily. On the other hand, Old World members of the subfamily belong to the loudest night singers of many forests. This disparity in their acoustic behavior has been explained by different hunting strategies of insectivorous bats in the New and the Old World, which in turn, shaped different defensive strategies among acoustic insects. Nearly all species of the Pseudophyllinae are arboreal or at least associated with tall vegetation. A notable exception is the genus *Callimenellus*, which is known to inhabit marine littoral rock crevices in Hong Kong. All species seem to be phytophagous although opportunistic insectivory has been observed in a few Central American species. Virtually all species are nocturnal.

Tettigoniinae. A large subfamily, with nearly 1000 described species. Most of the species of this subfamily occur in the temperate regions of the world, and only a handful of species are present in the tropics. Tettigoniinae achieved the greatest diversity in the regions of Europe, western North America, southern Africa, and Australia characterized by the Mediterranean type of vegetation. Many species have a large, shieldlike pronotum, hence the common name "shield-backed katydids." Females always have a long, swordlike ovipositor and lay eggs in soil or the stems of herbaceous plants. Males stridulate loudly and many species are active during the day. A few species of shield-backed katydids are agricultural pests, the best known being the Mormon cricket (*Anabrus simplex*) of the western United States. However, some species in China and Japan have been kept as pets for hundreds of years, and their pleasing calls have made a remarkable impact on the poetry and other arts of these countries.

Conocephalinae. The conehead katydids are a large, poorly defined, and possibly polyphyletic subfamily of Tettigoniidae, with more than 1100 described species worldwide. Many are characterized by a prominent fastigium of the vertex, forming a characteristic "horn" on the head. The diet of many of these katydids is restricted to grasses and their seeds, although quite a few species are predaceous. The conehead katydids are quite common in temperate regions of North America and Europe, although they reach their greatest diversity in the tropical areas of South America. Some species of Conocephalinae may become agricultural pests and a few have been known to form large, locust-like swarms (*Ruspolia*). Male investment in offspring can be significant in some neotropical conehead katydids, and an elaborate courtship behavior is common in such species.

Phasmodinae and Zaprochilinae. These two, closely related subfamilies are restricted to Australia. Phasmodinae contains only one genus, *Phasmodes*, with three species restricted in their distribution to Western Australia. These insects are a good example of convergent evolution, resembling walking sticks (Phasmodea) to an extraordinary degree. Both sexes are completely wingless, and females lack tibial tympana as well as thoracic auditory spiracles. The entire body is extremely elongate and thin, and the head is prognathous. A simple way of telling these insects apart from walking sticks is looking at the proportions of thoracic segments. In real walking sticks the prothorax is extremely elongated, whereas in *Phasmodes* it is the mesothorax that underwent such modification. These interesting katydids feed on both leaves and flowers of the heath habitats.

Zaprochilinae have similarly elongated bodies and the prognathous head, but all species have wings. They are strongly reduced in the genus *Kawanaphila* but fully developed in the remaining three genera of the subfamily. Males of all 28 described species produce short, ultrasonic calls. Several species of the subfamily have been extensively studied with regards to the parental investment of males, which is significant and may lead to courtship role reversal. All members of the subfamily feed primarily on pollen and nectar, and they play some role in pollination of the flowers on which they feed.

In addition to the above-described subfamilies, the family Tettigoniidae also includes the following subfamilies: Austrosaginae, Hetrodinae, Lipotactinae, Listroscelidinae, Meconematinae, Mecopodinae, Microtettigoniinae, Phyllophorinae, Saginae, and Tympanophorinae.

III. GRYLLOIDEA (CRICKETS AND MOLE CRICKETS)

A. Introduction

The superfamily Grylloidea includes more than 3700 described species placed in more than 520 genera and three families (Fig. 2). Crickets are small to medium size insects, the smallest species being only about 1.5 mm long (Myrmecophilidae) and the largest about 60 mm (Gryllotalpidae). The family is cosmopolitan in its distribution and some species occur even within the polar regions, albeit usually only in association with human dwellings. Most members of the subfamily are brown or black, very rarely green. The antennae are usually thin, threadlike and longer than the body (with the exception of certain burrowing forms). The head is usually large and almost globular, although it may be somewhat elongate and prognathous in some lineages (Oecanthinae, Gryllotalpinae). The pronotum is usually quadrate, rarely produced backward as to cover the wings (a condition common in katydids). The tarsi of the legs are three-segmented (superficially four-segmented in Oecanthinae), and the front legs are sometimes strongly modified for digging. The front tibia typically have well-developed bilateral tympanal organs. Hind legs are usually saltatorial but are occasionally short and not adapted for leaping. Wings are usually present but may be reduced to various degrees or totally absent (Myrmecophilidae). The tegmina are held flat on the dorsum and their anterior (lateral, in the resting position) margins are bent downward at an angle, forming a boxlike structure. The stridulatory apparatus of males is well-developed in most groups although it may be absent even in some fully winged taxa (some Pentacentrinae, Trigonidiinae, Eneopterinae). The left and right tegmina of males have similar venation, but stridulation is performed virtually always with the right tegmen over the left. The hindwings, when folded, often project far beyond the tegmina. Many taxa lack the hindwings and wing polymorphism is common. Male cerci are usually long, flexible, often superficially annulated, and never grasping. The ovipositor is usually long and needle-like, although it may be reduced or even absent in some crickets (Gryllotalpidae).

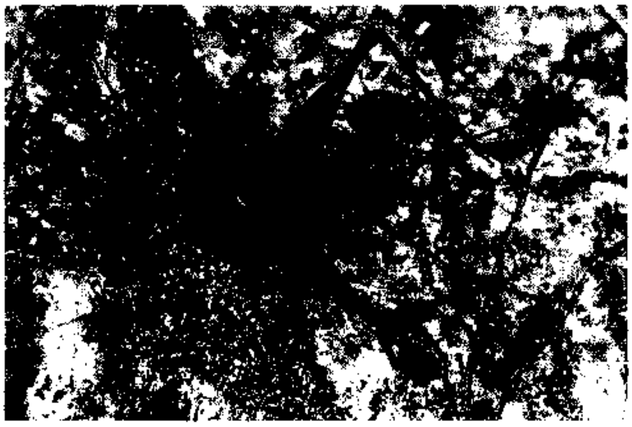


FIGURE 2 Grylloidea. *Aclodes* sp. from Costa Rica (Gryllidae, Phalangopsinae).

Most species of crickets are nocturnal or crepuscular. They occur in almost all terrestrial habitats, from treetops to underground burrows, and many species are associated with aquatic and semiaquatic environments (such forms are usually very good swimmers). The great majority of crickets are opportunistic feeders and few seem to be exclusively predaceous.

B. Major Lineages

1. Gryllotalpidae (Mole Crickets)

This interesting family includes five genera and 77 described species of usually large, robust crickets, distributed worldwide. All species of the family spend their lives in underground tunnels, excavated with their extremely modified, shovel-like front legs. The head is prognathous and somewhat elongated, and so is the pronotum. The antennae are short and somewhat thickened, an apparent adaptation for their subterranean lifestyle. The wings are usually fully developed, rarely absent. The stridulatory apparatus of males is simple, lacking the mirror. The females' tegminal venation resembles that of the males but lacks proper stridulatory organs. Cerci in both sexes are long and flexible, acting as an extra pair of antennae while the insects are moving backward in their underground corridors. Females lack an external ovipositor.

Mole crickets emerge from their tunnels only during courtship, although calling takes place underground. In order to broadcast their signals as far as possible, males build Y-shaped tunnels, the length and diameter of which are perfectly suited to amplify the dominant frequencies of their call. Females stridulate as well, but their call seems to have a territorial rather than courtship role. Eggs are often laid in special brooding chambers and actively cared for by the female. Newly hatched nymphs stay in the chamber for a few weeks, feeding on humus and tender rootlets protruding into the burrow.

Several species of mole crickets are serious crop pests, not only because they feed on roots of plants but also because they damage entire root systems while digging tunnels. A species of *Gryllotalpa* recently introduced accidentally to Australia has quickly become one of the major threats to the Australian golfing industry by damaging carefully cared for putting greens.

2. Gryllidae (True Crickets and Tree Crickets)

The family Gryllidae is the largest lineage of the superfamily, with nearly 3600 species, assigned to more than 500 genera in 17 subfamilies (there are several alternative systems of classification of crickets, and the number and status of higher categories varies among different authors). The body size of true crickets ranges from small (under 5 mm) to large (about 50 mm). Most members of the family are rather stout insects, with short, thick legs. Phalangopsinae are a notable exception, having extremely long and slender legs and other appendages, making them resemble large spiders rather than typical crickets. Delicate bodies with slender legs also characterize most tree crickets (Oecanthinae). The head is generally globular, except in tree crickets, which have an elongated, prognathous head. The pronotum is generally quadrate. The wings and stridulatory apparatus may be fully developed, reduced, or totally absent. Often, fully winged forms lack the stridulatory apparatus, while forms with greatly reduced tegmina can have a fully developed one. Species with well-developed stridulatory apparatus usually produce melodious, frequency and time-modulated calls, and many species have been kept as singing pets in Asia and Europe for many centuries. Rarely, the call is short and clicklike (*Eneoptera*). Courtship behavior is often complex and involves acoustic, tactile, and olfactory signals. The courtship song, produced only in the close proximity of the female, is quieter and of a different structure than the advertisement call. It is usually produced by rubbing only a part of the stridulatory file against the scraper. Some spider crickets (Phalangopsinae) additionally drum with their legs. During copulation, males produce spermatophores, which in some species contain a large spermatophylax. Males of some species allow the female to feed on different parts of their bodies. Males of tree crickets have thoracic glands, which the female licks during copulation, and females of some pygmy field crickets (Nemobiinae) feed on males tibial spines during copulation.

The ovipositor in true crickets is usually long, needle-like, sometimes laterally flattened and swordlike (Trigonidiinae) or reduced (*Brachytrupes*). Oviposition takes place in soil or plant tissues. Maternal care is rare but sometimes quite well developed. Females of *Anurogryllus muticus* excavate extensive subterranean burrows as a nursery for their eggs, which are aggressively protected against all intruders. Newly hatched nymphs stay with their mother, who feeds them with unfertilized eggs, produced by her for the sole purpose of feeding her nymphs.

a. Major Subfamilies

Gryllinae. This subfamily includes the true crickets, placed in more than 850 species and more than 100 genera. They are distributed worldwide, and a few species closely associated with human habitats are virtually cosmopolitan. Morphologically, Gryllinae are rather uniform, having stocky bodies, with a globular head and thick legs. Most have well-developed wings and a stridulatory apparatus. True crickets are some of the first singing insects to be heard in the spring in the temperate areas of Europe and North America, as they often overwinter as late nymphal instars and mature as soon as the temperature allows them to become active. North America has a particularly interesting fauna of true crickets, with several complexes of morphologically nearly identical, so called "cryptic species." Some of the species appear to be chronospecies, that is, very closely related species that avoid interbreeding by being sexually mature during different times of the year.

Most true crickets are ground dwellers and many are quite territorial, vigorously defending their burrows. In some species a certain percentage of the males (satellite males) consistently exhibit a kind of sexual parasitism by intercepting females attracted to the call of another male. Several species form large swarms, and some (e.g., *Gryllus bimaculatus* in Africa) can become serious agricultural pests. The house cricket (*Acheta domesticus*) is a cosmopolitan insect associated with human dwellings and when occurring in large numbers can cause damage to stored food and other material.

Brachytrupinae. This subfamily, often treated as a subset of the true crickets, includes about 20 genera and more than 220 species distributed mostly in the tropical and subtropical regions of the Old World, with relatively few species in the New World. The wings in some species are greatly reduced or absent and the female's ovipositor is reduced. Some of these crickets are very large and produce calls of an intensity directly proportional to their size. Some species are gregarious and live in small underground colonies consisting of adults and nymphs of different ages. Some species of the subfamily exhibit extraordinary maternal care of eggs and nymphs.

Phalangopsinae. The spider crickets are characterized by having a rather robust body supported, in most cases, by very long, thin legs. The antennae and palps in such forms are also extremely long and slender. Wings are frequently reduced or absent, and some winged species lack the stridulatory apparatus in the male. On the other hand, some species are loud singers, with a pleasing call, and males of several Asian species are greatly prized for their songs and are kept in special little cages as pets. The subfamily includes about 380 species in nearly 100 genera, achieving the greatest diversity in the tropics of Central and South America. Many species inhabit caves and rock crevices as well as spaces between buttress roots of trees in tropical forests. Others prefer tree trunks, leaf litter, and decaying wood.

Mogoplistinae. The scaly crickets are small, rarely exceeding 15 mm in length. Their name is derived from the fact that their entire body is covered with minute scales, reminiscent of the scales on the wings of butterflies and moths. Similar to these insects, the scales on the body of scaly crickets often form beautiful color patterns. Unlike butterfly wings, however, the wings of scaly crickets are not covered with scales, and are strongly reduced, often entirely concealed under a somewhat elongated pronotum. About 270 species have been described, placed in 26 genera. They are distributed worldwide but Australia and the Indo-Malaysian region seem to have the largest share of known species.

Oecanthinae. The tree crickets are rather unusual members of the family Gryllidae, having a somewhat elongated, prognathous head and a rather long pronotum. The wings can be either fully developed or completely absent. Winged forms have a very well developed stridulatory apparatus, capable of producing loud, pure sounds. Males of the genus *Oecanthus* enhance the range of their call by positioning themselves in a hole chewed out in a leaf, with their tegmina aligned during singing with the surface of the leaf. By doing this, they significantly increase the area from which the sound radiates, using the same principle on which speakers in radios and other musical appliances are built. Thanks to this technique, these small insects are capable of becoming the dominant singers in many environments.

Females of tree crickets lay eggs in the stems of plants, and by doing so may damage young trees in nurseries and orchards, thus becoming pests. On the other hand, many species feed on aphids, thus balancing the negative effect of their reproductive behavior.

Other subfamilies included in the family Gryllidae are Cachoplistinae, Eneopterinae, Euscyrtinae, Gryllomiminae, Itarinae, Malgasiinae, Nemobiinae, Pentacentrinae, Podoscirtinae, Pteroplistinae, Sclerogryllinae, and Trigoniidinae.

3. Myrmecophilidae (Ant-Loving Crickets)

This small family seems to be closely related to Mogoplistinae and probably does not deserve the status of a separate family. Only about 10 genera and less than 70 species are known, and all of these are closely associated with colonies of social insects, mostly ants, and less frequently termites. All species are minute, sometimes less than 1.5 mm, with an oval, wingless body and a small head with greatly reduced eyes. The hind femora are short and enormously enlarged.

Ant-loving crickets seem unable to live without a close association with ants or termites. Some species have a wide range of host species while others are restricted to only one host species. The nature of this association is unclear. At least one myrmecophilous species, *Myrmecophila manni* from North America, appears to be able to mimic ant behavior to an extent that allows it to be fed by workers of its host, *Formica obscuripes*. Another species, *M. oregonensis*, feeds on residues of the substance that lubricates ants' bodies, left on the walls of passageways. A few species of these crickets appear to be parthenogenetic. Their eggs are relatively large and their embryonic development is quite long, sometimes lasting over a year.

IV. GRYLLACRIDOIDEA (CAMEL CRICKETS, CAVE CRICKETS, JERUSALEM CRICKETS, RASPY CRICKETS)

A. Introduction

The taxonomy and phylogeny of this group of the Orthoptera seems to be in greater disarray than any other lineage of the order. Many different names have been applied to this group of insects, and frequently various subgroups have been elevated to the levels of superfamilies or even infraorders. As no modern phylogenetic study has yet been done on the Gryllacridoidea, for the sake of taxonomic stability this review follows classification proposed by Rentz (1996).

The superfamily Gryllacridoidea includes about 1500 species and more than 200 genera assigned to four families: the Cooloolidae (Cooloola monsters), the Stenopelmatidae (Jerusalem crickets and wetas), the Gryllacrididae (raspy crickets), and the Rhaphidophoridae (the cave and camel crickets) (Fig. 3). Members of this superfamily are characterized by the lack of the tegminal stridulatory apparatus even in fully winged species. Tegmina, if present, are soft and characteristically curved around the abdomen. The head is usually large and bullet-like, and some forms have greatly enlarged mandibles. The antennae tend to be very long and threadlike but in some subterranean forms (Cooloolidae) may be extremely shortened. Tibial auditory structures are usually absent, and the thoracic auditory spiracle is small and simple. Acoustic communication in this group is not as common as in other orthopterans and is restricted to femoral stridulation (Gryllacrididae) and abdominal drumming (Stenopelmatidae).

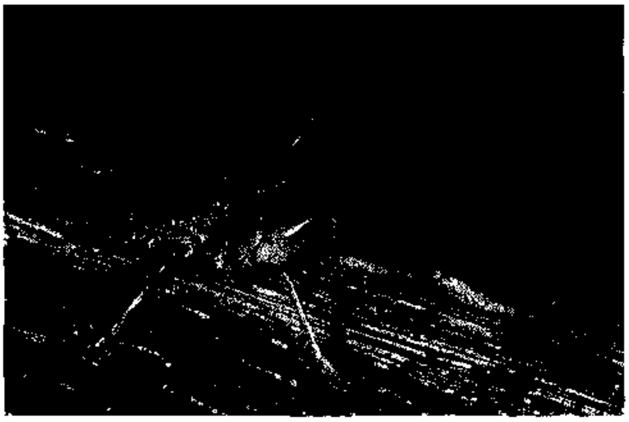


FIGURE 3 Gryllacridoidea. *Gryllacris* sp. from Costa Rica (Gryllacrididae). See also color insert, Volume 1.

Body size varies from about 10 mm to about 100 mm, and some species of New Zealand wetas are considered the heaviest insects in the world, reaching the weight of 70 g—nearly three times as much as an average house mouse. Coloration of species of Gryllacridoidea is rather drab, usually brown, black, or yellow, but never green.

The distribution of the superfamily is worldwide, with the exception of the Cooloola monsters, which are known only from northeastern Australia. At least one species, *Tachycines asynamorus*, has become virtually cosmopolitan thanks to its association with greenhouses and other human environments.

Nearly all species of Gryllacridoidea are nocturnal or crepuscular. Their habitats include treetops in tropical forests, savannas, and deserts, but a great proportion of species dwells in caves or underground burrows. Many species have a tendency to form large aggregations.

B. Major Lineages

1. Cooloolidae (Cooloola Monsters)

Some of the most peculiar insects belong to this small family, consisting of only one genus and four species. They are restricted in their distribution to central and northern Queensland in Australia and were discovered only about 20 years ago. Their appearance is so unusual that the first specimens ever collected were suspected of being fabricated as an entomological joke. The body of Cooloola monsters is so bulky that these insects, and especially the females, are almost incapable of walking on the surface of the ground. They spend their entire lives underground, digging through sand with their extremely thick and powerful legs. They are virtually blind and their very short antennae consist of only nine segments. Their mandibles are elongate and sharp, and the maxillae (the second part of mouthparts) bear two dagger-like processes, which probably act as piercing organs, a situation absolutely unique within Orthoptera. The wings are vestigial in males and usually completely absent in females. The internal anatomy of these insects is also unusual in the fact that the foregut extends backward into the posterior part of the abdomen.

Little is known about biology of Cooloola monsters. They appear to "sand swim" through sand and do not construct burrows. They feed probably on body fluids of other subterranean animals.

2. Stenopelmatidae (Jerusalem Crickets and Wetas)

This family includes about 230 described species, placed in more than 50 genera and six subfamilies. They are known primarily from North and South America, southern Africa, Madagascar, India, Australia, and New Zealand. The body of these insects is usually fairly large and bulky, and New Zealand wetas are among the largest living insects. The head is large and globular, with relatively short antennae. Legs are thick, powerful, and frequently well adapted for digging. In some species, the powerful hind legs can serve as a weapon, and these species are capable of delivering strong and painful kicks. Wings are usually absent and are rarely developed well enough for the insects to be able to fly. They do not have stridulatory organs on the tegmina, but many species communicate with elaborate percussive signals produced by drumming their abdomen against the substratum. The courtship behavior of Jerusalem crickets is a violent spectacle, involving a battle between the male and the female. If the female does not recognize the male as an acceptable partner, she devours him, thus preventing copulation. If the male is found to be acceptable, she allows him to mate and deliver to her a large spermatophore, and she attempts to eat him only afterward. This behavior may account for the rarity of male specimens of Jerusalem crickets in entomological collections. On the opposite end of the spectrum, males of some New Zealand wetas maintain harems of females in their refuges in holes and cracks in trees. Other males frequently try to take over the harems, and fights involving head-on clashing and gaping of their jaws are quite common among males.

Most species of Stenopelmatidae spend their lives on or close to the ground, with the exception of some Central and South American species (*Anabropsis*), which are often found in the canopy of the rain forest. Some North American species (*Cnemotettix*) produce silk with their mouthparts, which is used to line their burrows in coastal sand dunes. Many species are predaceous, but most seem to be opportunistic generalists. Jerusalem crickets have been reported to cause minor injuries to crops by gnawing on roots and parts of the plants close to the ground.

3. Gryllacrididae (Raspy Crickets and Leaf-Rolling Crickets)

This large family includes nearly 800 described species placed in almost 90 genera. It is quite likely, however, that these numbers represent only a fraction of the actual number of species, as the faunas of Gryllacrididae of Central and South America are virtually unknown. Most species of the family have been described from southeast Asia and Australia, but they are also known from Africa and the neotropics. North American fauna of Gryllacrididae is limited to only one species, *Camptonotus carolinensis*, and they are absent from Europe.

Species of Gryllacrididae are characterized by the globular head with often extremely long antennae, which in some species may exceed the body length four to five times. The wings are usually present but may be reduced or absent in some species. The ovipositor of females is long and thin, and in nymphs held curved backward over the body. This morphological modification may allow them to fit more easily into their burrows or rolled leaves. Leaf-rolling behavior is quite common in this family. All stages of some species, from the tiniest nymphs to adults, are capable of producing silk, which they use to bind shelters made of leaves, twigs, sand, and other material. Burrowing species living in hot, arid zones are known to insulate themselves from the heat by sewing shut the entrances to their burrows. The unusual subfamily Schizodactylinae (by some authors considered a separate family or even an infraorder) is unique in the peculiar modifications of their hindwings, which are coiled into tight, vertical spirals. Their feet possess long, lobelike processes, which help them run on loose sand.

Tegminal stridulation is absent in species of Gryllacrididae but many species produce raspy sound (hence the common name) by rubbing the inner side of the femora against the sides of the abdomen. This behavior does not seem to play any role in courtship but is rather a defensive strategy, since all life stages of both sexes are known to display it. All species of Gryllacrididae are nocturnal. They are probably unique in the insect world in possessing individually recognizable pheromones, which allows a cricket to return to its own burrow, distinct from one of its conspecifics, after a night of foraging. Most members of the family are predators and feed on small insects. Some are extremely agile and actively hunt for prey by rapidly running along branches, searching for insects.

4. Rhaphidophoridae (Cave Crickets and Camel Crickets)

Cave and camel crickets are represented by about 470 described species assigned to over 70 genera in seven subfamilies. They are distributed worldwide and at least one species is cosmopolitan. Species of this family are rather uniform in their appearance. All are completely wingless and have smooth, dorsally convex bodies. Most have long, spidery legs and similarly long and slender antennae and palps. The ovipositor is laterally flattened and often has distinctly serrated edges.

Virtually all species are associated with dark and humid habitats, such as caves and holes in the ground. Several species sometimes invade cellars and basements of houses, and one species (*Tachycines asynamorus*) is cosmopolitan in greenhouses and other manmade environments. Most species appear to be gregarious, and many individuals of different ages are found in a single colony. Many species are associated with burrows of larger animals, mostly rodents or foxes. They feed on a great variety of organic material, and some species are associated with bat colonies, feeding on guano and the bodies of dead bats. At least one species, *Pallidotettix nullarborensis* from Australia is known to cause death of unprotected chicks of nesting birds by swarming them and feeding on their tissues.

V. EUMASTACOIDEA (MONKEY GRASSHOPPERS OR EUMASTACIDS)

A. Introduction

The majority of members of this large superfamily occur in the tropics, although a few are found in temperate regions (Fig. 4). Nearly 1300 species have been described, and placed in about 300 genera in two families. As in all Caelifera, the antennae of eumastacids are very short, but unlike other members of the suborder, one of the last few segments of the antennae always bears a small tubercle of unknown function. The thorax can be short (Eumastacidae) to extremely long (Proscopidae), and in some groups it can have a distinct cuticular crest. The hind femora typically are long and slender, although in some forms may be greatly shortened and not adapted for leaping. In forms with long, leaping hind legs, they are characteristically splayed outward from the body when the insect is at rest. Wings are frequently reduced or absent, but when present they are very long, and characteristically widened toward the end of the body. The tegmina are frequently transparent or semitransparent. The abdominal auditory organ is absent in most species of the superfamily.

FIGURE 4 Eumastacoidea. *Homeomastax robertsi* from Costa Rica (Eumastacidae, Eumastacinae).

Little is known about the biology of members of this superfamily. None of the species appears to stridulate. Many species are arboreal, but some live on grasses in savanna-like environments and frequently have extremely elongated bodies, perfectly mimicking blades of grass (Australian Morabinae). Eggs are laid in batches in soil but, unlike those of other grasshoppers, they are never bound together by a foamy mass.

B. Major Lineages

1. Eumastacidae (Monkey Grasshoppers)

The larger of the two families in this superfamily, the Eumastacidae includes more than 1000 species in nearly 270 genera, distributed worldwide. Body size in this family is small (10 mm) to medium (45 mm). The pronotum is always short, and the front and middle legs are relatively short. Hind legs are usually saltatorial but in some grass-mimicking forms may be shortened and lacking the ability to leap. The male genitalic structures are very complex and of great value in taxonomic research. The ovipositor of females is short, with strongly serrated edges.

2. Proscopidae (False Walking Sticks or Proscopids)

Proscopids are restricted in their distribution to South America, barely reaching the southernmost part of the isthmus of Panama. About 300 species are known, placed in 29 genera. This family is readily distinguishable from other grasshoppers by their greatly elongate, twiglike appearance. Some species may reach a length of 165 mm, while the smallest are about 25 mm long. The head of these insects is unusually long and pear-shaped, with large eyes situated near its narrow tip. The thorax is extremely long and slender, and the front and middle legs are well separated. Femora of the hind legs are characteristically thickened at their bases and are not adapted for jumping. Virtually all species are completely wingless. Most species live on bushes and trees, and a few in savanna-type habitats. One species has occasionally been injurious to crops.

VI. PNEUMOROIDEA (BLADDER HOPPERS OR PNEUMORIDS)

This small group of grasshoppers includes some of the most unusual orthopterans living today (Fig. 5). The superfamily includes only 25 described species, placed in 13 genera of three families. The family Pneumoridae is restricted in its distribution to southern and eastern Africa, while families Tanaoceridae and Xyronotidae occur only in Mexico and southeastern United States.

The principal family, the Pneumoridae, includes large (50–100 mm) grasshoppers, with the wings either fully developed or somewhat reduced. The pronotum is greatly enlarged, forming a triangular helmet covering a large part of the body. The most unusual feature of the members of the family is the shape and size of the male's abdomen, which in most species is grossly inflated, creating a balloon-like, semitransparent resonator. The stridulatory apparatus consists of a dozen or so cuticular ridges on the third abdominal segment. Thanks to the abdominal resonating chamber, the sound made by the males is among the loudest sounds produced by any insect. Females are capable of producing sound as well but lack the abdominal stridulatory apparatus. Long winged species are sometimes called "flying gooseberries" as they are frequently attracted to lights. Males who have the misfortune of being attracted to campfires explode like balloons if they fall into the fire.

The remaining two families of Pneumoroidea include only eight species of small grasshoppers, living on shrubby vegetation in deserts of North America.

VII. ACRIDOIDEA (GRASSHOPPERS AND LOCUSTS)

A. Introduction

Grasshoppers and locusts form the largest lineage of Orthoptera, with more than 8500 species in nearly 1800 genera assigned to 10 families (Fig. 6). They are distributed worldwide, and although they are most diverse in tropical and subtropical regions of the world, they also occur in cold, subpolar regions and high in the alpine zones of the mountains. The body size of these insects ranges from less than 10 mm to over 120 mm, with a wingspan of 250 mm. The body form and the head shape are very diverse, ranging from short and bulky to extremely elongate and slender. The antennae are always short and relatively thick, but often modified to form comb- or clublike structures. The hind legs are always distinctly thickened toward the base and always adapted for leaping. The wings vary from fully developed to completely lacking. Fully winged forms have relatively very narrow tegmina that are not widened toward their apices. Stridulatory organs are present only in some families and usually consist of a modified vein on the tegmen, against which the inner side of the hind femur is rubbed. The abdomen usually has lateral auditory structures, even in silent and wingless species. The female's ovipositor is always very short, with the upper and lower portions (valves) distinctly divergent

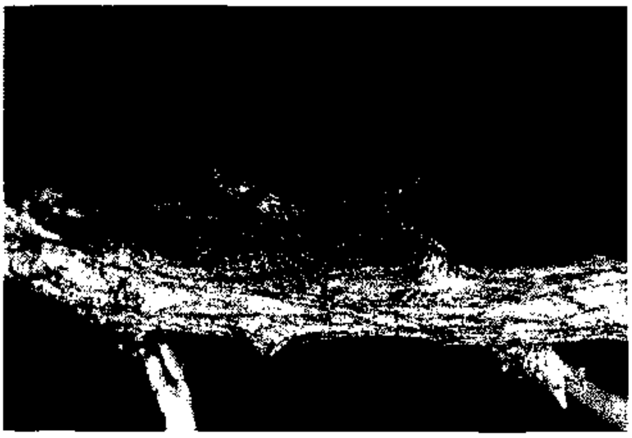


FIGURE 5 Pneumoroidea. *Mohavacris* sp. from the southwestern United States (Tanaoceridae).

FIGURE 6 Acridoidea. *Chortoicetes terminifera* from Australia (Acrididae, Oedipodinae). See also color insert, Volume 1.

at the apex. Eggs are laid primarily in the soil, and are always protected with a foamy mass produced with special glands in female's abdomen.

Most grasshoppers are diurnal, although nocturnal species are also known. They inhabit nearly all terrestrial habitats and a few species are aquatic (Paulinidae). Open steppes and savannas of warmer regions of the world have particularly rich faunas of grasshoppers, but tropical rain forests have their share of unique forms too. Grasshoppers are also relatively diverse in temperate zones, with some species crossing the polar circles. A number of alpine species are adapted to activity in very low temperatures. Some of these forms exhibit the interesting behavior of changing their body color depending on air temperature. In cold weather, they become much darker, and thus more likely to absorb sun rays (*Kosciuscola*).

Virtually all species of grasshoppers are exclusively phytophagous. Opportunistic cannibalism has been observed in few rare instances, but only under crowded conditions. Most grasshoppers are polyphagous, feeding on a number of different plant species, although most tend to have preferences toward certain species of plants. Some grasshopper species are monophagous, feeding on only one species of plant throughout their life. Such species tend to feed on plants rich in toxic secondary compounds, and in doing so become unpalatable to potential predators themselves. They often advertise their toxicity with bright, aposematic coloration. Some grasshoppers (African Pyrgomorphidae) can even be harmful to humans, if ingested.

A few grasshopper species, mostly of the subfamilies Cyrtacathacridinae, Oedipodinae, Calliptaminae, and Melanoplinae, are very serious agricultural pests and cause enormous losses to crops every year. They form unbelievably huge swarms, which upon descending on fields can devour the entire annual crop within minutes. Single swarms of the African plague locust *Schistocerca gregaria* may fan out over hundreds of square miles and consist of 50 billion (50,000,000,000) individuals, weighing about 70,000 tons. Densities may reach 200 million locusts per square mile (= 500 million/km^2). In most species of locusts, this kind of behavior is seasonal, and some have two distinct morphological forms, the so-called solitary and gregarious phases. Sometimes the differences between the solitary and gregarious phases are so dramatic that they have been described as different species. The solitary forms occur in low densities, and usually have shorter wings and cryptic coloration. Under certain conditions, especially following rains, the densities of these grasshoppers increase dramatically, prompting great physiological, morphological, and behavioral transformations. Their coloration changes, usually to black and yellow, and they exhibit a strong tendency to form tightly packed groups that start marching across the land. Upon reaching the adulthood and the ability to fly, they take off in huge swarms, usually following the prevailing winds of a cold front that could lead them to areas of fresh vegetation.

Controlling massive outbreaks of grasshoppers is difficult and many different chemical and biological agents have been used. As one can imagine, controlling a cloud of insects stretching across hundreds of miles is a nearly impossible task. Therefore, the most effort is directed toward controlling the young, freshly emerged bands of hoppers or destroying eggs laid in areas that are likely to originate massive outbreaks. Fungal pathogens have been quite successful in combating locusts, but the survival of fungal spores in arid conditions, where many locust species live, is limited.

B. Major Lineages

1. Pamphagidae (Earth Hoppers or Pamphagids)

Members of this family are large, robust grasshoppers, seldom smaller than 30 mm and often up to 90 mm long. More than 560 species in more than 90 genera have been described, mostly from arid regions of Asia and Africa, with a few representatives in southern Europe. Sexual dimorphism is often marked, with males in general being significantly smaller and often fully winged, as opposed to completely wingless females. The surface of the body is often very rough, and many species blend superbly among rocks and pebbles. Some southern African species (*Trachypetrella*) have been sometimes dubbed "living rocks."

Defensive stridulation is common in both nymphs and adults of many species, and it is achieved by various mechanisms (often more than one stridulatory mechanism is found in a single species). The typical advertisement stridulation is rare but, if present, the sound produced by males by rubbing their hind legs against the tegmina is exceptionally loud (*Loboscelianа*).

2. Pyrgomorphidae (Bush Hoppers)

More than 550 species are known in this family, placed in more than 150 genera. They are distributed mostly in the tropics of the Old World, with only a few species known from Mexico and Central American countries.

The body is usually large, often reaching 90 mm. The pronotum is extremely variable in form, sometimes adorned with granules or spinelike processes. Pyrgomorphidae have no stridulatory organs but some make rustling noises with their wings when alarmed. Many members of this family have warning (aposematic) coloration, a characteristic frequently associated with their toxic properties. Some species eject protective froth or irritating fluids as defense mechanisms. Quite a few species are poisonous. Some species are gregarious, especially during their early nymphal stages. A few have been known to cause serious damage to crops in Africa (*Zonocerus*).

3. Romaleidae (Lubber Grasshoppers)

This family includes more than 470 species in more than 100 genera, distributed mostly in North and South America, with very few species in Asia and eastern Africa. In the United States the best-known representative of this group is the Florida grasshopper (*Romalea microptera*), a large and colorful grasshopper frequently used to study insect internal anatomy and external morphology.

Romaleidae are large grasshoppers, usually strikingly colored. Species of *Tropidacris* and *Titanacris* are some of the largest orthopterans, with wingspan reaching 250 mm. The pronotum in the members of this family is often strongly keeled or crested. Wings, if present, have dense venation, and the hindwings often have a unique stridulatory mechanism formed by numerous fine teeth on one of the veins. Sound is produced by rubbing this vein against the underside of the tegmen. In addition to stridulation, some species produce a hissing sound by expelling air through their thoracic spiracles. Species of this family have little economic significance, although a few species can defoliate trees on tropical plantations.

4. Paulinidae (Aquatic Grasshoppers)

Members of this small family are unique among other grasshoppers in their nearly fully aquatic lifestyle. They are capable of skating on the surface of the water, diving, and swimming, and their eggs are laid on stems of submerged aquatic plants. The hind tibiae are flattened and widened at their apices, forming effective swimming paddles. Only two genera with five species belong to this family, and their original distribution was restricted to South America. Recently, one species (*Paulinia acuminata*) has been deliberately introduced to Africa in an effort to control the noxious aquatic plant *Salvinia*. This is probably the only example thus far of a truly beneficial species of grasshopper.

5. Acrididae (True Grasshoppers and Locusts)

This is the largest family of Acridoidea, with nearly 6760 species in more than 1350 genera and 27 subfamilies. They occur in all regions of the world, and in nearly all terrestrial or semiaquatic habitats. The body size ranges from less than 5 mm to about 100 mm. Body form and shape are extremely variable, from robust and stocky to extremely slender and elongate. Wings can be fully developed, reduced, or entirely absent. Wing polymorphism is common in some groups (Gomphocerinae). Abdominal tympanal organs are generally present, absent only in some wingless forms. Many species stridulate by rubbing their hind femora against modified veins on the tegmina. Sometimes an expanded area on the tegmen forms a "speculum," which acts as a resonator.

All species of locusts and most other economically important grasshoppers belong in this family. Also, most grasshoppers common in temperate zones of the world belong here, and nearly the entire grasshopper fauna of Australia is the result of a massive adaptive radiation within one of the lineages of this family (Catantopinae).

VIII. TETRIGOIDEA (PYGMY GRASSHOPPERS OR GROUSE LOCUSTS)

This interesting lineage of the Orthoptera includes nearly 1200 species in about 230 genera, placed in one family, Tetrigidae (some authors divide this superorder into two families: Tetrigidae and Batrachideidae) (Fig. 7). Most species of pygmy grasshoppers are tropical but quite a few occur in temperate zones, often at very high elevations. Their body size is generally small, seldom exceeding 20 mm. The pronotum is always very large,

FIGURE 7 Tetrigoidea. An unidentified species from Costa Rica (Tetrigidae).

covering the entire body and often extending far beyond the end of the abdomen. Sometimes the pronotum is vertically expanded, forming a large crest or a leaflike lobe. A few species display a polymorphism in the development of the pronotum. Wings are usually present but may be reduced or absent in some forms. The tegmina are always reduced to small, oval lobes, with greatly reduced venation. The hindwings are well-developed, fanlike, usually completely concealed under the pronotum. The hind femora are very stout and all tarsi lack the arolium between claws. Members of Tetrigoidea have neither stridulatory organs nor the abdominal auditory tympana.

Most species are associated with moist or semiaquatic habitats. Many can swim very well and dive when threatened. Some species are arboreal and many live on bark of trees in tropical forests. Pygmy grasshoppers feed on a variety of small plants, such as mosses, as well as on lichens and algae. They have little economic importance although some may cause some damage in rice plantations.

IX. Tridactyloidea (False Mole Crickets and Sandgropers)

This superfamily includes about 230 species in 18 genera divided among three families (Fig. 8). Most species are small, 4 to 15 mm in length, although some sandgropers (Cylindrachetidae) can reach the length of about 40 mm. Families Tridactylidae and Ripipterigidae have characteristically modified hind tibiae, which bear a pair of long, apical flaps, while the tarsi are strongly reduced. Their tegmina, if present, are short and leathery and sometimes have a row of stridulatory pegs. The Tridactylidae are distributed worldwide whereas the Ripipterigidae are restricted to Central and South America. They inhabit wet and semiaquatic habitats, and some of them make dense networks of shallow burrows in the mud.

The Cylindrachetidae are restricted to Australia and New Guinea, and one species is known from Patagonia. The body is strongly elongated and completely wingless. The pronotum is greatly elongated and cylindrical, with its lateral edges strongly curved downward, nearly meeting on the underside of the thorax. The front legs are modified for digging, and the second and third pairs are shortened. Sandgropers spend their entire lives underground in sandy soils. They are sometimes confused with the larvae of click beetles (Elateridae) as they occasionally tunnel into stems of plants.

FIGURE 8 Tridactyloidea. *Neotridactylus apicalis* from the northeastern United States (Tridactylidae).

See Also the Following Articles

INSECTS, OVERVIEW

Bibliography

Gangwere, S. K., *et al.* (eds.) (1997). The bionomics of grasshoppers, katydids and their kin. CAB International.

Kevan, D. K. McE. (1982). Orthoptera. In *Synopsis and Classification of Living Organisms, Vol.* 2 (S. P. Parker, ed.), pp. 352–379. McGraw-Hill, New York.

Otte, D. (1981). *The North American Grasshoppers. Vol. I.* Harvard University Press, Cambridge.

Otte, D. (1984). *The North American Grasshoppers. Vol. II.* Harvard University Press, Cambridge.

Otte, D. (1994). The crickets of Hawaii. *Publications on Orthopteran Diversity.* The Orthopterists' Society at the Academy of Natural Sciences, Philadelphia.

Rentz, D. C. F. (1979). Comments on the classification of the orthopteran family Tettigoniidae, with a key to subfamilies and description of two new subfamilies. *Australian Journal of Zology* 27, 991–1013.

Rentz, D. C. F. (1986). *Grasshopper Country: The Abundant Orthopteroid Insect Fauna Of Australia.* University of New South Wales Press, Sydney.

Rentz, D. C. F., and John, B. (1990). Studies in Australian Gryllacrididae: Taxonomy, Biology, Ecology and Cytology. *Invertebrate Taxonomy* 3, 1053–1210.

Uvarov, B. P. (1966). *Grasshoppers and Locusts: A Handbook of General Acridology. Vol. I.* Cambridge Published for the Anti-Locust Research Centre, London.

Uvarov, B. P. (1977). *Grasshoppers and Locusts: A Handbook of General Acridology. Vol.* 2. Centre for Overseas Pest Research, London.

Vickery, V. R., and McE. D. K. (1985). The grasshoppers, crickets, and related insects of Canada and adjacent regions. *The insects and arachnids of Canada, Part 14.*

GRAZING, EFFECTS OF

Mark Hay and Cynthia Kicklighter
Georgia Institute of Technology

GLOSSARY

associational refuge When a potential prey escapes or deters a consumer by associating with another organism that interferes with the ability of the consumer to locate or attack the prey.

bottom-up effects When physical parameters (such as nutrients) allow increased primary productivity and the effects of this cascade up through higher trophic levels. A simple example involves nutrient additions to a lake resulting in increased phytoplankton growth, this increasing resource of food allowing increased numbers of herbivorous zooplankton, and the abundance of zooplankton leading to increased densities of zooplankton-eating fishes.

co-evolution Joint evolution of two (or more) interacting species, each of which evolves in response to selection imposed by the other.

competitive release The expansion of a species-realized niche that is associated with the absence or removal of competition from other species.

grazer A consumer that removes only a part of each prey it attacks and thus rarely kills a prey in the short term. The term is commonly applied to animals that eat plants (herbivores) but can also be argued to apply to nonlethal microbial pathogens and to animals that cause tissue loss, but not death, when they feed from colonial animals such as sponges.

resource partitioning When organisms differentially utilize resources such as food, space, or nutrients.

secondary metabolite An unusual compound that does not play a role in primary metabolism (e.g., production and storage of energy). These metabolites were initially argued to be waste products but have often been shown to defend the producer from consumers, pathogens, or competitors.

top-down effects When feeding by higher trophic-level organisms has cascading effects on lower trophic levels. A simple example is when fishes (predators) consume zooplankton (herbivores) and, by lowering the numbers of these herbivores, allow increased densities of photosynthetic phytoplankton (primary producers).

GRAZING in the traditional sense of the term describes the actions of animals that consume parts of plants without causing the death of the plant; e.g., the con-

Encyclopedia of Biodiversity, Volume 3

sumption of grass plants by domestic cattle. By extension, the term applies more generally to any feeding process that involves the partial consumption of the prey organism so as to bring about tissue loss but not death; e.g., the consumption of coral reef organisms by sea urchins and certain fishes. In either sense of the term, the processs of grazing has profound implications for the structure and dynamics of the given community.

I. INTRODUCTION

Herbivores play a major role in determining plant biodiversity by strongly affecting distribution, abundance, and community organization. Because plants are the major primary producers and often constitute the major biogenic structural complexity around which other community members are organized, effects of herbivores on plants often have strong indirect effects on other trophic levels and on the biodiversity of the ecosystem as a whole.

Because almost all energy on Earth is fixed via plant photosynthesis, it is an evolutionary mandate that higher trophic levels will need to consume plants and that successful plants will need to resist, tolerate, or escape herbivores, pathogens, and other natural enemies. Humans are dependent on plants for food, clothing, shelter, fuels, pharmaceuticals, and other necessities; thus, we often complete with other consumers (e.g., insect herbivores and pathogens of crop plants) for desirable plant resources. Due to this continuing struggle to grow and harvest plant resources, plant–herbivore interactions are one of the more intensively studied areas in ecology, evolution, and agriculture. Research on this topic provides a rich resource for evaluating the effects of grazing on biodiversity. In numerous instances, examples from both marine and terrestrial studies show similar patterns. However, because of the more rapid growth of marine plants, the more intense grazing by marine herbivores, and the greater ease of manipulating seaweeds versus trees or shrubs, some of the clearest examples investigating the mechanisms of how herbivores affect plants are from marine communities.

In subtidal marine communities such as coral reefs or kelp beds, it is not uncommon for herbivores to be one of the primary forces determining the distribution and abundance of plants and often the species composition and diversity of the entire community. For example, herbivory by fishes or sea urchins on coral reefs keeps reefs largely devoid of macroalgae and allows corals to flourish by reducing competition from the more rapidly growing seaweeds. When these herbivores severely declined on the Caribbean island of Jamaica due to a combination of overfishing and urchin disease, coral reefs suffered severe overgrowth by seaweeds, and coral cover declined from more than 50% to less than 5% throughout this entire island nation (Fig. 1). Thus, removing herbivores from these habitats converted species-rich coral reefs into a completely different community dominated by a limited number of seaweeds.

Equally dramatic changes occur in temperate systems. Numerous ecological and paleontological studies have shown that subtidal communities in the temperate eastern Pacific shift from sea urchin-grazed barrens to lush kelp beds depending on the presence or absence of sea otters, which selectively forage on herbivorous sea urchins, reducing grazer biomass and allowing kelps to flourish. If otters are removed from kelp communities due to hunting or predation, urchin densities increase, urchins drive kelps to local extinction, and fishes, kelp-associated invertebrates, and seals or sea lions that feed on kelp bed fishes also decline dramatically. Thus, in the tropical example, grazers increase diversity by removing seaweeds and allowing corals to produce the biogenic structure that enhances reef biodiversity, whereas in the temperate system grazers directly remove the major biogenic structure (kelps) and cause a decline in overall biodiversity.

Herbivores can also cause major alterations in terrestrial communities. When myxomatosis eliminated rabbits in Britain, many grasslands reverted to scrub woodlands. Similarly, it is argued that grazing elephants

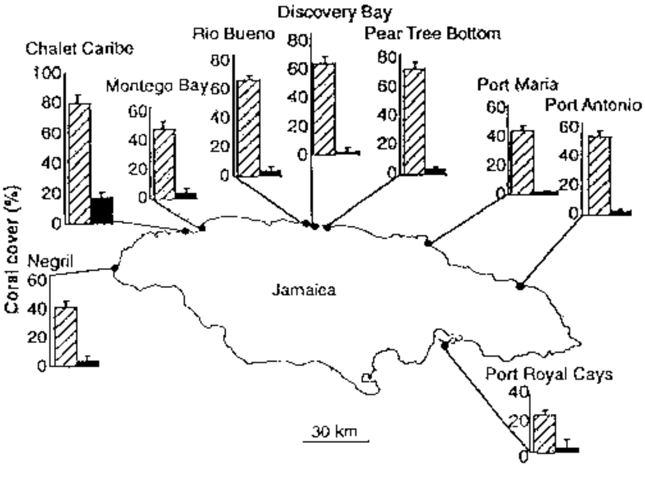


FIGURE 1 Large-scale changes in percentage cover of live coral at fore-reef sites along >300 km of the Jamaican coastline between 1977 (hatched bars) and 1993 (solid bars). For the island as a whole, coral cover decreased from 50% to <5% (reproduced with permission from Hughes, 1994, Catastrophies, phase shifts, and large-scale degradation of a caribbean coral reef. *Science* **265**, 1547–1550. Copyright 1994 American Association for the Advancement of Science).

damage trees and change closed woodlands or thickets into grassy savannas.

The previously mentioned instances are among the most dramatic examples of herbivore effects, but they serve to show the effects that grazing can have on biodiversity and community structure and to point out that there should be strong selection for plant traits that deter feeding by herbivores.

II. GRAZERS AS DISTURBANCE REGIMES

Charles Darwin noted that one grass species tended to dominate and exclude other species in unmown areas of his lawn, but that numerous species coexisted in areas that were periodically mown. He recognized that grazers might play a similar role in diversifying plant communities by preventing a few rapidly growing species from competitively excluding others. In the late 1970s, Joe Connell formalized this relationship when he demonstrated for both tropical rain forests and coral reefs that intermediate frequencies or intensities of disturbance could lead to coexistence of a larger number of species. When disturbances were severe and frequent, only a limited number of hardy species could occupy a habitat. When disturbances were too infrequent or weak to remove competitive dominants, superior competitors tended to dominate. However, when disturbances occurred at intermediate levels, they removed enough of the competitively superior species to allow invasions by others but were not so severe as to prevent the occurrence of less hardy species. Thus, these intermediate levels of disturbance promoted the maintenance of greater biodiversity within a given community type. One of the most omnipresent forms of disturbance to plant communities is grazing, and it can play the diversifying role suggested by Connell. However, the effects that grazers have on plant biodiversity may also depend on the productivity potential of the environment. A recent review of how grazing affected plant diversity used data from a diverse range of marine, terrestrial, and fresh-water communities and found that grazing almost always increased plant diversity in environments characterized as having high nutrients (and presumably high potential growth rates), but that grazing often decreased diversity or had no effect on diversity when the environment was nutrient poor. These findings can be argued to mean that there are important grazer–nutrient interactions that need to be factored into the intermediate disturbance model. Alternatively, these findings simply mean that a given level of disturbance is "high" if it occurs in a habitat in which growth rates (and thus the ability to recover from the disturbance) are low, but it is "intermediate" if it occurs in a habitat in which rapid growth allows quick recovery.

As an example of intermediate grazing diversifying plant communities, work by Hixon and Brostoff (Fig. 2) demonstrated that fish grazing rates on open areas of coral reefs were so intense that only the most resistant (coralline algae, which are red algae that resemble a thin calcified paint applied to a rock) or most tolerant (small filamentous algae that have basal sections embedded in the coral and that grow very rapidly) seaweeds could persist. In contrast, when fishes were excluded by cages, substrates were dominated by a few species of competitively superior macroalgae. However, in defended territories in which damselfishes drove most other herbivores away, bite rates on the substrate were intermediate between those experienced on open versus caged substrates, and at these intermediate rates of herbivory there were many more species of seaweeds. Similar patterns have been documented in Mediterranean grasslands in which plant species diversity is maximized if livestock densities are maintained at intermediate levels.

III. DIRECT VERSUS INDIRECT EFFECTS

Grazers may indirectly affect biodiversity via trampling, tunneling, seed dispersal, nutrient regeneration, and selection for plant traits that have cascading effects on other processes. As herbivorous prairie dogs dig their burrows, they create patches of newly disturbed ground that can be colonized by plant species that are rare in undisturbed habitats. Their activities also create mosaics of vegetation structure and nutrient status that result in both buffalo and antelope differentially using centers or edges of prairie dog colonies versus areas remote from colonies. Prairie dogs also clip inedible bushes to low heights so that the bushes do not provide ambush sites for predators or obscure the prairie dog's field of view. Lowering bush height decreases competitive effects and allows increased coexistence of other plants. As animals graze, they release nutrient-rich urine and feces back into the environment; this release affects nutrient dynamics, the spatial patchiness of plant production, and thus plant diversity. Such effects occur in habitats ranging from open-ocean gyres to African savannas and livestock rangelands.

Initial feeding on plants can also produce a cascade of indirect effects on habitat parameters that then affect use by other species. For example, when infestations of beetles feed on goldenrod, they not only affect their

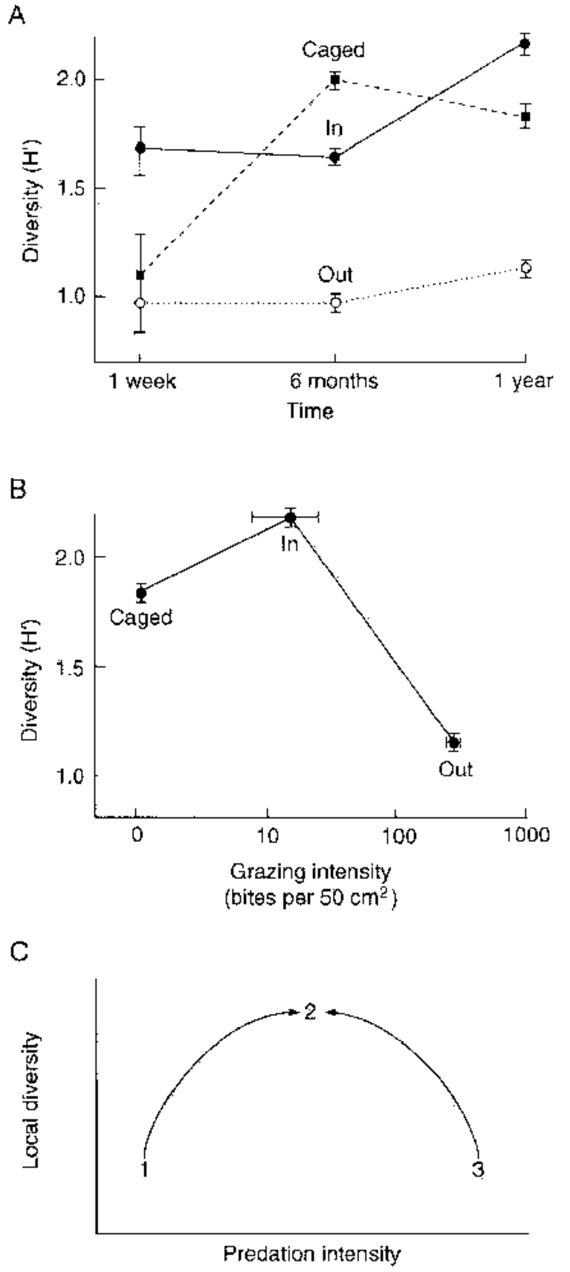


FIGURE 2 (A) Diversity, as measured by the Shannon–Weiner index H′, of algae on settling plates as a function of exposure time to three treatments: caged, inside damselfish territories (In), and outside territories (Out). (B) Diversity after 1 year (same data as in A) as a function of the intensity of fish grazing. (C) Graph of the intermediate disturbance hypothesis showing that consumers can enhance local diversity by increasing predation intensity from point 1 to point 2 or by decreasing overall predation intensity from point 3 to point 2 (reproduced with permission from Hixon and Brostoff, 1983, Damselfish as keystone species in reverse: intermediate disturbance and diversity of reef algae. *Science* 220, 511–513. Copyright 1983 American Association for the Advancement of Science).

target plant but also their activities alter light penetration to the ground, soil water content, and soil nitrate content. Following these changes in the physical habitat, areas near attacked goldenrod are invaded by other plant species, increasing local biodiversity.

IV. THE ROLES OF SPATIAL AND TEMPORAL SCALES OF GRAZING

Many investigations of herbivore effects on diversity have concentrated on the maintenance of within-habitat diversity due to selective feeding on dominant plants or on the effects of small-scale disturbances that create patches and produce numerous successional states within a single community. In contrast to these studies, several investigations from marine systems show that species richness of seaweeds is often maintained by mosaics of herbivore impact that are relatively predictable in space and persistent in time. Many seaweeds are predictably found only within certain physical habitats. This restricted distribution was initially assumed to be due to physiological limitations or to fine-scale resource partitioning along physical gradients; however, experimental studies demonstrated that many of these restrictions were herbivore induced and did not result from habitat partitioning based on competitive interactions or differing physiological requirements. When spatial variations in herbivory are experimentally reduced, habitats that differ dramatically in seaweed composition often become more uniform, usually resulting in a lowering of total species richness due to the loss of between-habitat diversity. In most cases, species that dominate following herbivore removal are palatable species that exclude, or significantly reduce the abundance of, herbivore-resistant plants. This is consistent with the hypothesis that the evolution of anti-herbivore defenses can be achieved only by diverting energy and nutrients from other needs. Thus, defenses appear to be costly, and in the absence of herbivory less defended individuals or species will have higher fitness than more heavily defended individuals or species. This relationship is depicted graphically in Fig. 3. When herbivores are rare, plant communities are dominated by competitively superior, but palatable, plants. When herbivores are common, defense is favored over rapid growth and competitive ability. Under these conditions, communities become dominated by unpalatable plants that depend on the herbivores to remove better competitors. If herbivores selectively graze competitive dominants, they tend to free resources for other species and increase biodiversity. However, if herbivores selectively graze

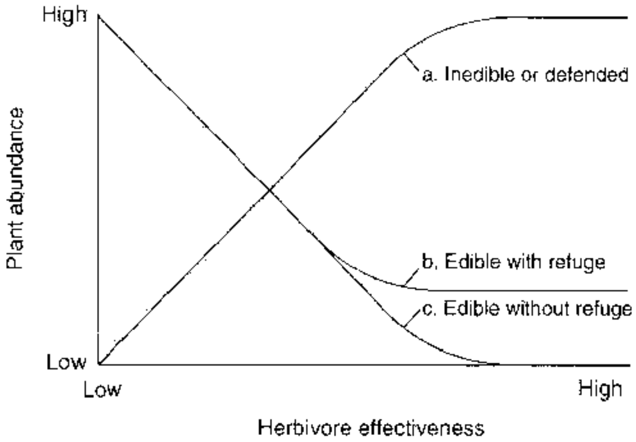


FIGURE 3 The hypothesized relationship between plant abundance and herbivore effectiveness for palatable and unpalatable plants. a, Inedible or defended plants; b, palatable plants with a refuge; c, palatable plants without a refuge. Spatial patchiness of herbivory within a habitat allows multiple points on the *x*-axis to exist within the same habitat, thus increasing local biodiversity (reproduced with permission from *Annual Review of Ecology and Systematics*, Volume 12, © 1981, by Annual Reviews *www.annualreviews.org*).

competitively inferior species, they may hasten local extinction of poorer competitors and lower biodiversity. This latter scenario appears to be relatively uncommon.

In the following sections, we discuss how herbivores affect plant diversity over a range of spatial scales. Patterns occurring over a scale of hundreds of kilometers are called geographic patterns. Patterns that occur over a scale of several meters to many kilometers and involve habitats that clearly differ in physical traits are called between-habitat patterns. Patterns occurring over only a few meters and within areas that appear relatively homogeneous are called within-habitat patterns. Patterns on a scale of centimeters or millimeters are called microhabitat patterns.

A. Geographic Patterns

Large-scale patterns in plant distribution are usually assumed to be due to physical parameters such as temperature. However, numerous correlative studies suggest that some portion of geographic constraints on both specific species (or phenotypes within species) and morphological types may be due to geographic variance in herbivory. Because rates of herbivory are higher on tropical than on temperate reefs, one might predict that tropical plants would be better defended than temperate plants. There are few direct tests, but a recent study demonstrated that seaweeds from tropical reefs were significantly less palatable and significantly better defended chemically than were similar seaweeds from temperate reefs. Similarly, studies of many species of salt-marsh plants along the coast of the eastern United States show that individuals from southern regions are less palatable to a range of common herbivores than are individuals of the same species from northern regions. These patterns suggest that herbivore-generated processes shown in Fig. 3 could function on geographic and local scales and could enhance global biodiversity by creating differing selective regimes across these latitudinal scales.

Because tropical herbivores encounter more chemically defended prey than do temperate herbivores, one might expect that tropical herbivores would be more tolerant of plant chemical defenses than would temperate herbivores. This contention has not been adequately tested, but initial tests of a few tropical versus temperate fishes and sea urchins are supportive of this hypothesis.

B. Between-Habitat Patterns

Near-shore reefs often support more seaweed biomass and a different community of seaweeds than offshore reefs. Although both physical parameters and herbivores differ between these habitats, field experiments demonstrate that these differences in seaweed community structure are determined to a significant degree by differential herbivore activity. Early studies in the Caribbean demonstrated that seagrass beds, unstructured sand plains, and shallow unstructured reef flats supported different plant communities than did immediately adjacent reefs. Although the different physical regimes of these habitats appeared to be an adequate explanation for the differing seaweed communities, caging and transplant studies demonstrated that removal of herbivores allowed a more homogeneous distribution of seaweeds across these different habitats and resulted in lowered biodiversity of the reef system as a whole. In several studies, seaweed species typical of shallow reef flats invaded and dominated deeper areas on the reef once herbivores were removed. For example, when large fishes were excluded from parts of a coral reef in Belize, large macrophytes that were typical of nearby reef flats (but normally absent from the subtidal reef) invaded in large numbers and significantly decreased the abundance of corals and herbivore-resistant seaweeds during a 10-week period. Once barriers were removed, this 10 weeks of algal growth was totally consumed in less than 48 hr. A common response in herbivore exclusion experiments has been for palatable seaweeds to competitively suppress or exclude less palatable (and presumably better defended) seaweeds, again suggesting a tradeoff between defense and competitive ability.

C. Within-Habitat Patterns

By aggressively excluding other herbivores from their algal mats, territorial damselfishes generate patches of intermediate grazing intensity in which algal species richness, evenness, and diversity are increased relative to either areas that are available to all grazers or areas from which grazers are excluded by cages. Several species that are normally excluded from unprotected areas due to grazing are found only within these mats. Thus, these territorial herbivores create distinct biotic patches on an otherwise more homogeneous background, raising species richness of the community as a whole.

In a similar manner, predators can strongly impact the spatial patterns of habitat use by herbivores and indirectly create mosaics of grazing intensity. In streams, algal-eating minnows avoid both shallow areas where they are susceptible to terrestrial predators and areas near larger predatory fishes. In both cases, these behaviors create patches of increased algal mass separated by heavily grazed habitats with minimal algal mass. Similar mosaics are created in terrestrial habitats when desert rodents forage more heavily under bushes than they do in more open areas where they are more exposed to predators. Within grasslands, excavating mammals, such as badgers, prairie dogs, and pocket gophers, provide newly bared soils for fugitive plants that are unable to successfully invade undisturbed plots. These microhabitat mosaics produce a series of successional states within the community and enhance plant diversity.

D. Microhabitat Patterns

In marine hard-substrate communities, small cracks and crevices serve as microhabitat escapes from some herbivores. Under these conditions, herbivory on exposed surfaces may select for defended seaweeds, whereas selection favors competitive ability in cracks and indentations. Work in experimental microcosms has shown that algal diversity decreases with increasing numbers of herbivorous fishes when substrate is smooth, but that species number changes little with grazer densities if the substrate is more topographically complex. Thus, the presence of microhabitat escapes provided spatial refuges for palatable species and prevented their exclusion from the system.

Some types of plants can generate microhabitat escapes. There are many instances in agriculture and in the plant–insect literature in which interspersion of host plants with non-host species interferes with the ability of insect herbivores to find or effectively utilize their host. Spiny desert plants sometimes function as "nurse plants" where less well-defended juvenile plants escape grazing vertebrates until they become large enough to escape because of their size. Plankton communities also exhibit these associational refuges; herbivorous copepods may quit feeding on palatable species of phytoplankton when they are mixed with adequate densities of chemically noxious species.

The effects of associational refuges on species richness have been addressed experimentally in several marine communities. Numerous seaweeds that are commonly driven to local extinction by grazers persist in herbivore-rich communities by growing on or beneath their herbivore-resistant competitors. The brown seaweed *Stypopodium zonale* produces cytotoxic compounds that deter feeding by Caribbean reef fishes and urchins. Numerous species of seaweeds are significantly more common near the base of *Stypopodium* plants than several centimeters away. When the chemically deterrent plant is removed, these less defended species are rapidly removed by herbivores, decreasing local species diversity. When plastic mimics of *Stypopodium* are placed in the field, they provide a partial refuge for palatable species but they are less effective than the real plants, suggesting that associational refuges are generated in part by the physical presence of a nonfood plant but that the plant's chemical repugnance makes the associational refuge more effective.

In temperate communities, palatable seaweeds can reduce losses to herbivores by growing on or near unpalatable seaweeds. Growing in close association with these unpalatable competitors drastically depresses the growth of palatable species, but the associational benefits, in terms of reduced herbivory, can more than offset this competitive cost. Thus, palatable species can be dependent on their unpalatable competitors to produce spatial refuges from herbivory and prevent their exclusion from the community. Both field and mesocosm studies indicate that removal of common unpalatable competitors can cause extinction, rather than competitive release, of associated competitors that are more palatable. These associational refuges were initially interpreted as arising from simple visual crypsis, but more detailed investigations suggest that chemistry plays a significant role.

As a final example, sulfuric acid can constitute up to 18% of the dry mass of the brown alga *Desmarestia*; this concentration is sufficient to dissolve barnacles from coastal rocks when this alga is deposited in the intertidal by waves. In Chilean kelp beds heavily grazed by sea urchins, the palatable kelp *Macrocystis* cannot successfully colonize unless it invades an area encircled by *Desmarestia* plants, which appear to act as acid brooms that prevent urchins from entering the area.

The associational escapes discussed previously are opportunistic rather than coevolved. As such, they may be used by many organisms in a wide variety of situations. There are, however, more intimate associational refuges or defenses that may be coevolved. Many microbes are predictably associated with specific species of macro-organisms. Because of the broad ability of microbes to produce bioactive secondary metabolites, many host organisms could be coevolved with certain microbes because the microbes produce compounds that defend the host from natural enemies. For example, certain fungi infect host grasses, produce toxins, and by doing so make the grasses much more resistant to herbivores. Similarly, marine cyanobacteria grow in host sponges and produce bioactive secondary metabolites that can protect the host. There are also instances in which marine bacteria that are specialized to certain host surfaces produce metabolites that chemically defend their host from microbial pathogens. In such cases, pathogens or consumers may be selecting for specialized microbial associates that defend their host. Such associations could increase microbial diversity and the diversity of macro-organisms by allowing hosts to persist in new habitats, thus facilitating both the evolution and the retention of increased species diversity.

V. HOST CHEMISTRY AS A PROMOTER OF CONSUMER DIVERSITY

On some tropical reefs, fishes have been reported to bite the bottom in excess of 150,000 times/m^2/day. In these areas, small herbivores such as amphipods, polychaetes, and crabs (collectively called mesograzers because of their size) would live short lives if they occupied plants that were preferred by fishes. Selection should therefore favor sedentary mesograzers that live on and eat seaweeds that are chemically repellent to fishes. Patterns supporting this notion have been documented for divergent types of mesograzers in several of the world's oceans. In the temperate Atlantic, some herbivorous amphipods and polychaetes live in tubes which they attach to the seaweeds they consume. These mesograzers selectively live on and feed from brown algae that are chemically defended from fishes. By living in association with these seaweeds, which are seldom visited by fishes, they lower their susceptibility to predation.

The hypothesis that sedentary mesograzers minimize predation by specializing on toxic hosts has been tested more broadly using (i) a specialist Caribbean amphipod that eats, and builds a mobile domicile from, a chemically defended alga; (ii) several species of crabs and sea slugs from both the Caribbean and tropical Pacific that each live on and feed from only one species of chemically defended alga; (iii) non-herbivorous amphipods that live on and build homes from only one species of chemically noxious seaweed; (iv) a decorator crab that minimizes predation by selectively decorating with a seaweed that is chemically repellent to fishes; and (v) contrasts between the palatability and susceptibility to predation or parasitism of specialist versus generalist herbivorous insects. In all these cases, predation on, or palatability of, the mesograzers was reduced as a consequence of their association with chemically noxious hosts. Additionally, mesograzers were generally stimulated or unaffected by plant compounds that deterred feeding by larger herbivores or predators.

There are numerous examples indicating that seaweeds or sessile invertebrates that evolve effective defenses against reef fishes may become evolutionary targets for specialized mesograzers that can escape or deter their own consumers by evolving a resistance to these compounds and living on, feeding from, and in some cases morphologically mimicking or sequestering defensive compounds from their toxic hosts. Thus, once hosts effectively deter common fish predators, these hosts can serve as valuable spatial escapes from predation for small consumers that specialize on these noxious hosts. As in many of the earlier examples, this creates spatial mosaics of selection within otherwise more uniform habitats and allows for the evolution and retention of more species within the system.

VI. THE POTENTIAL IMPORTANCE OF PALEO-PATTERNS

Plant–herbivore interactions occurring today are the result of both modern ecological forces and an evolutionary history that has been operating for many millions of years. Major changes in plants or in herbivores can select for fundamental and often cascading changes in plant–herbivore interactions. For example, in modern African communities, grazing by elephants causes repeated deforestation of savannas, prevents larger plants from replacing grasses, and facilitates retention of other herbivore species that require open-grazing habitats. It has been hypothesized that loss of elephants and other megaherbivores from Europe and the Americas at the end of the Pleistocene may explain the dramatic loss of other grazing mammals in these regions. It can be argued that elephant grazing prevented open savannas from converting to woodlands and that grazing by rhinoceros and hippopotamus prevented short grasslands from converting to less nutritious tall grasslands. The loss of these large grazers would have caused mosaics of parklike woodlands and grasslands to convert to more homogeneous forests and less productive tall grasslands. The loss of the mosaic vegetation structure is predicted to have caused the loss of numerous smaller herbivores that depended on spatial patchiness and the availability of nutrient-rich vegetation. This hypothesis is consistent with the variability in extinction seen among Africa, Eurasia, and the Americas and with the absence of extinctions at previous glacial–interglacial transitions.

The mass extinction of megaherbivores at the end of the Cretaceous also coincided with a dramatic change in the structure and species composition of the plant community. Most of the dominant ferns and cycads went extinct, but angiosperms and gymnosperms flourished. The decreased herbivory following the mass extinction should have allowed vegetation to become much denser, potentially increasing competition among plants. This change is hypothesized to have lead to the evolution of larger seeds as a way of better provisioning juvenile plants. These larger seeds and the need for dispersal may have led to fruits for attracting animal dispersers. These abundant fruits and seeds are thought to have favored the evolution of small, fruit-eating birds and mammals that aided seed dispersal.

The cause–effect relationships hypothesized to have driven the paleo-changes discussed previously must be viewed with caution, given the general inability to rigorously test hypotheses associated with evolutionary time. However, it is clear that major changes in vegetational structure often coincide with extinctions of large herbivores and visa versa. There is also broadly accepted support for the notion that there has been an evolutionary arms race between vascular plants and their insect herbivores—the plants evolving chemical defenses against the insects, the insects responding by evolving resistance to some of these defenses and then tending to specialize on plants with those chemical signatures, and the cycle then repeating. It is hypothesized that this diverse range of plant chemical defenses serves as a rich and heterogeneous resource gradient that has promoted the dramatic diversification of herbivorous insects.

It is tempting to assume that coincident changes in herbivores and plants are caused by reciprocal and evolved responses. However, such "co-evolved" responses may be difficult to differentiate from fortuitous preadaptations. One marine study with an especially complete fossil record provides an example. Encrusting coralline algae are highly calcified crusts that cover hard substrates and fossilize so well that grazing scars attributable to specific types of herbivores (i.e., gastropods, sea urchins, or fishes) are clearly preserved on their surfaces. These algae have left a long and abundant fossil record. Using fossilized corallines and information on present-day interactions between corallines and reef herbivores, one can convincingly argue that the evolution of herbivorous fishes (especially parrotfishes) dramatically increased herbivory on coral reefs and fundamentally changed the selective regimes affecting seaweeds on reefs and reef community organization in general. Coincident with these changes in grazing pressure, coralline crusts with numerous traits that minimized the effects of herbivorous fishes radiated dramatically while their assumed parent taxon (the solenopores, which lacked these traits) diminished in importance and eventually went extinct. From these coincident changes, it seems reasonable to assume that the escalation of grazing drove the evolution of corallines and selected for the numerous morphological traits that minimized the effects of herbivores on corallines. However, the exceptionally complete fossil record for the corallines allows this assumption to be tested. Virtually all the morphological traits that appear to be "adaptations" to grazing were present in the corallines before the evolution of macroherbivores and hundreds of millions of years before the evolution of parrotfishes. Thus, grazing fishes may well have driven the solenopores extinct and may have changed reef ecology to favor the expansion of the corallines, but the corallines were fortuitously preadapted for these changes—they did not adapt in response to them. Such examples should encourage caution in ecological interpretation of the fossil record.

VII. USING GRAZERS TO MANAGE AND RESTORE ECOSYSTEMS

There is a long history of using grazers to maintain or restore ecosystem structure, function, or diversity. Both productivity and plant species composition can be affected by manipulating the duration, intensity, and timing of grazing. Herbivore manipulations have been used to increase species diversity, to increase the aesthetic value of habitats, and to preserve endangered species. Most grazers being used as management tools come from terrestrial grasslands, but there are indications that this practice might also be useful in a wide variety of habitats ranging from forests and heathlands to coral reefs and lakes. In the following sections, we discuss some examples by habitat type.

A. Grasslands

Grazers are crucial to the maintenance of grasslands because without them there is an accumulation of plant litter that sequesters nutrients, physically limits vegetative growth, and interferes with seedling establishment. In a 10-year experiment focused on chalk grasslands in The Netherlands, sheep grazing helped rebuild species-rich grasslands that had been degraded due to agricultural use and heavy fertilization. When grazed and ungrazed plots were compared, the grazed plots had about 50% more species and the relative mass of forbes was increased threefold. Studies of other grassland systems and grazers have produced similar results. For example, sheep grazing can be used to reclaim heathland from woodland by opening up vegetation, repressing growth of scrub, promoting low-growing plants, and encouraging growth of dwarf shrubs. Experiments with rabbits and bison show similar results. In both cases, when the grazers were removed from grasslands, plant diversity declined.

Although grazing increased biodiversity in the previously mentioned studies, several of these management efforts were designed to have this result, thus potentially providing a biased impression of how grazing will affect plant species richness. Under some conditions, grazers decrease biodiversity. Effects are likely to vary as a function of the evolutionary history of the plants and herbivores, the types of habitats investigated, the levels of grazing employed, and when and for how long these grazing regimes are maintained. As an example of the variable effects that grazing can have on producer diversity, a recent study reviewed 44 comparisons of plant species richness under low versus high grazing pressure in nutrient-rich versus nutrient-poor ecosystems. High rates of grazing reduced plant species richness in 100% of the nutrient-poor situations. In contrast, high rates of grazing increased plant diversity in 56% of the high-nutrient contrasts; diversity was unaffected by grazing in 36% of these contrasts and decreased by grazing in 8% of the contrasts.

Manipulation of herbivores can also be used to alter habitat traits, such as productivity, rates of nutrient cycling, and nutrient, water, and organic levels in the soil. In general, herbivores often increase organic breakdown and the mineralization of potentially limiting nutrients such as nitrogen and phosphorous. In the Serengeti, the aboveground productivity of moderately grazed plots is stimulated to about twofold greater than the productivity of ungrazed plots. In addition, ungrazed grass stands are senescent, whereas those grazed by large herbivores produce younger and more palatable shoots.

Although mowing can crudely substitute for some grazer effects, grazing will generally produce greater diversity due to spatial mosaics produced by patchy grazing (versus even defoliation by mowing) and localized trampling. Selectivity and patchiness of grazing will also be affected by herbivore morphology and behavior. For example, teeth size and mouthpart morphology are important in determining the degree of selectively exhibited by grazers. For example, cattle are less able to graze individual plants or plant parts than are sheep, goats, or horses.

Fire has also been used as a surrogate for grazing or, in conjunction with grazing, to manage plant community structure and species composition. This is especially true in attempts to prevent exotics and woody species from altering grasslands. Although fire may be a necessary disturbance in some systems, such as tallgrass prairie, fire alone is often not sufficient for restoration of biodiversity. After 9 years, in annually burned tallgrass prairie with nitrogen- addition, species diversity was 48% lower than it was 5 years before burning started and 66% lower than in unburned plots. In contrast, mowed, annually burned, nitrogen-addition plots had more than twice the species diversity than similar unmowed plots. Therefore, mowing prevented a decrease in species diversity that would have occurred with only fire and nutrient addition. When bison grazing replaced mowing, plots that were grazed and burned had the highest species diversity of all plots. Similarly, in a Mediterranean grassland study, cattle grazing led to an increase in species diversity, whereas burning resulted in no significant difference.

B. Forests

Many tropical forest trees produce large fruits that either fall near the parent tree or are dispersed by birds, primates, and other large mammals to sites remote from the adult tree. Dispersal is so critical for some of these species that their seeds require gut passage through dispersal agents before they can germinate. Because of the reliance on dispersers and the heavy predation on seeds near the parent canopy, it has been suggested that seed predators and dispersers are critical for maintaining tree species diversity in tropical forests. By selectively preying on seeds that fall near parent trees, specialized seed predators may decrease the survivorship of seeds near parents, preventing the occurrence of monospecific patches of forest trees and facilitating a more even and diverse assemblage of species. Because larger animals disperse the seeds of more than half of woody plant species, they play a critical role in removing seeds from sites of high predation and in introducing new species into different patches within a forest. The hunting, poaching, and habitat modifications that have left many forests depauperate in these animal dispersers have been suggested to be a cause of lowered plant diversity in tropical forests. If this is correct, then effective management of tropical forests may require the replenishment of large-bodied seed dispersers.

Livestock introduced to serve as seed dispersers can partially replace native dispersers that disappeared during the Pleistocene megafaunal extinction. The feeding activities of these introduced species can increase the range of some plants that produce large, fleshy fruits. In a lowland deciduous forest in Costa Rica, introduced horses and cattle feed on the fruits of jicaro (*Crescentia alata*) and disperse their seeds. In areas with livestock, jicaro are common. In areas in which livestock are absent, jicaro is relatively rare and occurs primarily in small, spatially restricted patches.

Many seed dispersers subsist exclusively on fruit for at least part of each year. This suggests that removal of seed dispersers by anthropogenic activities could have several repercussions. First, decreased seed dispersal could eventually lead to a decrease in the number of mature seed-producing parent plants. This would in turn lead to a decrease of important wildlife food resources, which could then lead to even lower numbers of seed dispersers. This situation can be especially critical if a particular tree species is pivotal in maintaining the health of dispersers during food-limiting seasons. For example, *Casearia* in tropical rain forests produce rich fruit during seasons of relatively low fruit abundance and likely supply much of the diet of local fruit-eating dispersers. In addition, this species supports frugivorous birds that are very important dispersers to many other tree species during seasons when fruit is more abundant. The disappearance of animals that feed on *Casearia* fruits, leading to decreased recruitment of *Casearia*, could have far-reaching effects on the forest community. This scenario is thought to have occurred with jicaro. Frugivorous bats also feed on jicaro, and it is thought that a decrease in jicaro following the loss of Pleistocene seed dispersers led to a decrease in bats and then to a corresponding decrease in other fruit-producing plant species on which bats feed.

C. Lakes

Attempts to regulate the community composition of lakes have produced a mature science in which the separate and interactive effects of both physiochemical and biotic interactions are reasonably well understood and used in novel ways to manage lake ecosystems. Success in these systems has been achieved by fusing an understanding of biotic processes such as competition and predation (traditionally studied by population and community ecologists) with an understanding of the role of physiochemical processes (traditionally studied by limnologists and ecosystem ecologists). Investigations by Steven Carpenter and colleagues have been especially important in stimulating this field.

Problems of harmful algal blooms, fish kills, and general eutrophication of lakes have become increasingly apparent. Initial attempts to explain these problems, and lake productivity in general, as a function of nutrient levels revealed that nutrients could vary considerably in lakes with similar biotic communities and that nutrient levels (also known as bottom-up effects) explained only approximately 50% of the variability in lake productivity. This finding prompted lake ecologists to investigate the effects of trophic cascades (or top-down effects) in structuring these ecosystems. By understanding and manipulating both the bottom-up effects of nutrients and the top-down effects of consumers, lake ecologist have been remarkably successful at altering biotic communities and fundamental processes (e.g., productivity and nutrient cycling) occurring in lakes.

Because lake communities are structured by interactions between physiochemical conditions and biotic processes, initial attempts to manage lakes by managing nutrient input met with only variable success. When the traditional focus on nutrients was merged with top-down manipulation of trophic structure, greater success was achieved in the control of lake phytoplankton populations and productivity. Mechanisms for this are as follows: As piscivore biomass increases, their feeding causes a decline in the biomass of small fishes that feed on herbivorous zooplankton; with the decline of these planktivorous fishes, the biomass of large herbivorous zooplankton increases and causes a decline in phytoplankton biomass and an increase in water clarity (Fig.

4). In all these instances, productivity of each group is maximized at intermediate levels of piscivore biomass. This system of trophic cascades allows managers to affect planktivore, herbivore, or phytoplankton biomass and productivity via manipulation of larger fishes (which can be done using fish additions, removals, or common fisheries management practices).

Whole lake manipulations have shown that piscivorous fish populations can regulate planktivorous fish populations in both North American and European lakes. For example, in an experiment by Carpenter and Kitchell, one lake had a high bass (piscivorous fish) population and a small minnow (zooplanktivorous fish) population, whereas another had a low bass population and a high minnow population. Ninety percent of the large bass population was placed in the bass-deficient lake, and 90% of the large minnow population was placed in the minnow-deficient lake. Where bass were added and minnows removed, the size of herbivorous zooplankton increased due to decreased predation. This led to decreases in chlorophyll concentration and primary production. In the opposite experiment, trophic interactions ultimately led to increases in chlorophyll concentration and primary productivity.

The interaction between trophic cascades and nutrient addition has also been addressed. In whole lake experiments in which phosphorous was added to lakes with a high or low density of piscivorous fishes, only a high density of piscivorous fishes was effective in maintaining low levels of palatable phytoplankton. The added nutrients increased the biomass of herbivores and colonial blue-green (unpalatable) algae but not the biomass of palatable algae. Grazers can control chlorophyll levels at phosphate loading significantly higher than loads that cause eutrophication. However, grazer control of undesirable blue-green algae appears to fail at phosphate loading rates lower than those used in this manipulation.

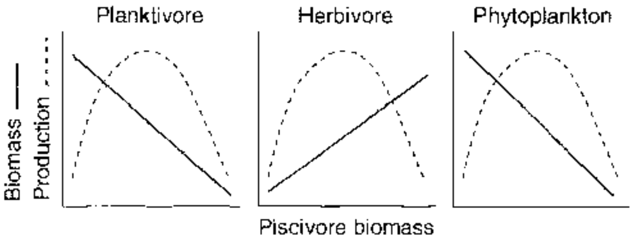


FIGURE 4 Effects of piscivore (= bass) biomass on biomass (solid line) and productivity (dashed line) of vertebrate zooplanktivores, large herbivores, and phytoplankton (reproduced with permission from Carpenter *et al.*, 1985, Interactions and lake productivity. *Bioscience* 35, 634–639 © 1985 American Institute of Biological Sciences).

D. Coral Reefs

Feeding by large generalist herbivores, such as fishes and sea urchins, is critical for maintaining the coral-generated topographic complexity that helps maintain biodiversity on tropical coral reefs. If these herbivores are removed, reef corals are competitively excluded by seaweeds. For example, overharvesting led to severely depleted populations of herbivorous fishes on coral reef in Jamaica. However, the effects of overfishing were not fully realized for decades because the lowered numbers of predatory and competing herbivorous fishes allowed a build-up of herbivorous sea urchins that largely compensated for the declining fish grazing. When the urchin population was severely reduced by disease, macroalgal cover increased from 4 to 92%, severely overgrowing corals and inhibiting coral recruitment. The few grazers left on the reef (small parrotfishes and surgeonfishes) were not enough to control the algal growth, which has persisted for more than a decade. From scenarios such as this, it is evident that a higher density and diversity of grazers are required to decrease algal cover and allow coral recovery.

Manipulating these herbivores would therefore be a logical way to manage critical biotic processes on coral reefs. Because of the open nature of marine populations (i.e., the long-distance larval dispersal typical of most reef herbivores means that populations must be managed regionally rather than locally), this is a more challenging proposition than in the lake or grassland systems discussed previously. However, there is considerable evidence that degrading reefs recover, or degrade more slowly, when they are protected from fishing than when fishing is allowed to continue. The increased health of the protected reefs is correlated with an increase in herbivorous fishes. Managing reefs to facilitate targeted herbivores that are especially important in removing reef macrophytes is a promising tool for aiding coral reef recovery. This more directed approach is currently being initiated but is not yet a proven tool for coral reef management.

VIII. SUMMARY

Coexistence of potential competitors within diverse plant communities has often been explained as a result of fine-scale resource partitioning. This is assumed to have resulted from an evolutionary history of competitive encounters and is usually inferred from shifts in abundance that occur between habitats with different physical characteristics. Both marine and terrestrial plants show predictable patterns of distribution that can be interpreted as this type of habitat partitioning. However, manipulative experiments in marine systems, and to a lesser extent some terrestrial communities, show that many plants are habitat generalists and that both between- and within-habitat patterns of distribution are often controlled by herbivores. Spatial patterns in herbivore effectiveness create a mosaic of habitats that differ in the degree to which they favor poorly defended but competitively superior species versus well-defended but competitively inferior ones. If herbivory is decreased, the mosaic nature of the habitat is reduced and many well-defended species are excluded by competition. This pattern occurs across a large range of spatial scales and can explain between-habitat differences in plant communities that occur hundreds of meters apart, within-habitat differences that occur only a few meters apart, and microhabitat differences that occur on a scale of centimeters or millimeters. A portion of the microhabitat pattern is created by the plants. For example, consumption of palatable plants can be significantly reduced when they grow on, or are intermixed with, unpalatable plants. These associational refuges increase species richness by allowing palatable forms to invade grazed areas after unpalatable forms establish and create microsites of reduced herbivory. At all of the scales examined, herbivore activities that promote spatial mosaics of herbivore effectiveness, or that differentially affect dominant plants, help to maintain higher biodiversity.

Foraging by herbivores is often constrained by factors (e.g., predation) that do not directly affect the physiological performance of primary producers. For plants, this differential use of space by herbivores often produces a spatial mosaic of selective regimes in an area that would otherwise be treated as one uniform habitat; this interaction commonly results in elevated species richness. The strong effects of herbivores on plants, and the strong effects of plants on community structure in general, mean that manipulations of plant–herbivore interactions can often be used as an effective management tool for promoting the maintenance of biodiversity or ecosystem function.

See Also the Following Articles

CATTLE, SHEEP, AND GOATS, ECOLOGICAL ROLE OF • COEVOLUTION • DISTURBANCE, MECHANISM OF • PLANT-ANIMAL INTERACTIONS • REEF ECOSYSTEMS • RESOURCE PARTITIONING • SEAGRASSES • SPECIES COEXISTENCE

Bibliography

Carpenter, S. R., and Kitchell, J. F. (1988). Consumer control of lake productivity. *BioScience* 38, 764–869.

Carpenter, S. R., Kitchell, J. F., and Hodgson, J. R. (1985). Interactions and lake productivity. *BioScience* 35, 634–639.

Hay, M. E. (1997). The ecology and evolution of seaweed–herbivore interactions on coral reefs. *Coral Reefs* 16, S67–S76.

Hay, M. E., and Fenical, W. (1996). Chemical ecology and marine biodiversity: Insights and products from the sea. *Oceanography* 9, 10–20.

Hixon, M. A., and Brostoff, W. N. (1983). Damselfish as keystone species in reverse: Intermediate disturbance and diversity of reef algae. *Science* 220, 511–513.

Hixon, M. A., and Brostoff, W. N. (1996). Succession and herbivory: Effects of differential fish grazing on Hawaiian coral-reef algae. *Ecol. Monogr.* 66, 67–90.

Hughes, T. P. (1994). Catastrophes, phase shifts, and large-scale degradation of a Caribbean coral reef. *Science* 265, 1547–1550.

Huntley, N. (1991). Herbivores and the dynamics of communities and ecosystems. *Annu. Rev. Ecol. Syst.* 22, 477–503.

Pacala, S. W., and Crawley, M. J. (1992). Herbivores and plant diversity. *Am. Nat.* 140, 243–260.

Proulx, M., and Mazumder, A. (1998). Reversal of grazing impact on plant species richness in nutrient poor vs nutrient rich ecosystems. *Ecology* 79, 2581–2592.

Schultz, E. D., and Mooney, H. A. (Eds.) (1993). *Biodiversity and Ecosystem Function*. Springer-Verlag, Berlin.

GREENHOUSE EFFECT

Jennifer A. Dunne and John Harte
University of California, Berkeley

GLOSSARY

aerosols Microscopic airborne particles.

albedo The fraction of light hitting a surface that is reflected.

anthropogenic Resulting from human activities.

climate sensitivity Long-term change in global mean surface temperature following a doubling of equivalent carbon dioxide (CO_2) concentration in the atmosphere.

Earth energy balance Average balancing of incoming solar energy by outgoing terrestrial radiation for Earth as a whole.

equivalent CO_2 concentration Concentration of carbon dioxide that would cause the same amount of radiative forcing as a mixture of carbon dioxide and other greenhouse gases.

feedback Change in a system component that triggers effects that eventually change the original component again. Feedbacks can be positive (self-reinforcing) or negative (self-dampening).

greenhouse gases Atmospheric gases that can absorb and reradiate infrared radiation.

radiative forcing Measure used to express and compare the potential of climate change factors to perturb the Earth energy balance, reported in watts per square meter ($W\ m^{-2}$). A positive radiative forcing tends to warm the Earth's surface and a negative radiative forcing tends to cool the surface.

THE "GREENHOUSE EFFECT" REFERS TO THE PROCESS by which infrared radiation-absorbing gases in Earth's atmosphere trap heat and thus influence climate. This article gives an overview of the anthropogenic loading of greenhouse gases into the atmosphere and associated effects on recent and future climate change, summarizes feedback effects, and describes potential and current impacts of climate change on biodiversity.

I. INTRODUCTION

Earth's climate, from daily weather events to glacial and interglacial cycles, is driven by the amount of radiation received from the sun and how that radiation is distributed throughout the global Earth–atmosphere system. The atmospheric greenhouse effect acts as an important factor in establishing a temperature that is hospitable for life. The basic mechanism is simple and was first detailed by the Swedish physicist Svante August Arrhenius in 1896. Light from the sun largely penetrates the atmosphere and is absorbed at the planetary surface. There, it is converted from energy in the form of light to energy in the form of heat (longwave infrared radiation). As the surface temperature rises because of this heat, Earth radiates more and more heat back out to space, thereby maintaining an energy balance. Several gases in the atmosphere, referred to as "greenhouse gases," absorb some of the heat emitted from Earth's surface and reradiate it back toward the surface, causing the temperature to rise. Without this naturally occurring greenhouse effect, Earth's average surface temperature would be −19°C, about 33°C colder than it is today. The term "greenhouse effect," though popular, is a misnomer because the warming effect of glass greenhouses is due primarily to suppression of convection, not trapping of infrared radiation.

II. ENHANCED GREENHOUSE EFFECT

The most important naturally occurring greenhouse gases are water vapor, carbon dioxide (CO_2), methane (CH_4), nitrous oxide (N_2O), and ozone (O_3). Although water vapor plays the biggest role in generating the natural greenhouse effect, anthropogenic emission of the other gases, along with artificially produced chlorofluorocarbons (CFCs), are most important in generating an enhanced greenhouse effect. Detailed instrument data show that concentrations of these gases have been increasing since preindustrial times (~1750) (Table I), particularly in recent decades, largely due to human industrial, agricultural, and urbanization activities. As the concentrations of greenhouse gases increase in the atmosphere, they continue to trap and reradiate more and more heat, resulting in rising surface temperature and other climatic changes.

Increases in CO_2 account for about 65% of the current direct positive radiative forcing due to anthropogenic loading of greenhouse gases (Fig. 1). The atmospheric CO_2 concentration has increased 30% since preindustrial times, as a result of increasing emissions from fossil fuel combustion, land conversion, and cement production, and is continuing to increase by 0.4% per year (Fig. 2). If future emissions of CO_2 are maintained at 1994 levels, its atmospheric concentration will be close to double the preindustrial level by the end of the twenty-first century. In the absence of strong emissions controls, given increasing global energy and resource consumption, CO_2 concentrations may double by 2040 and will continue to increase dramatically (Fig. 3). Concentrations of other greenhouse gases, particularly methane and nitrous oxide, are also expected to rise, resulting in an earlier doubling of the equivalent CO_2 concentration. Greenhouse gases tend to remain in the atmosphere for many years (see Table I) and consequently are well mixed. They continue to affect the climate long after initial emissions and later stabilization of atmospheric concentrations.

Aerosols can alter the climate by changing atmospheric albedo. These fine particles absorb and reflect solar radiation and alter cloud properties. Sulfate aerosols from fossil fuel emissions and smelting tend to have a negative effect on radiative forcing and thus cool the climate. Current estimates of direct radiative forcing are -0.5 W m^{-2} due to aerosols compared to 2.45 W m^{-2} due to greenhouse gases (see Fig. 1). Unlike greenhouse gases, aerosols are very short-lived in the atmosphere and therefore are not well mixed and respond rapidly to changes in emissions.

III. CLIMATIC CONSEQUENCES: GLOBAL WARMING

A. Past Climate Change

Scientists use deep ice cores drilled from glaciers in Antarctica, Greenland, and South America to examine ancient climate and atmospheric trends over the last several hundred thousand years. The data from these ice cores reveal a strong correlation between temperature and CO_2 concentrations (see Fig. 3) and suggest that a doubling of CO_2 has been historically associated with a 3 to 4°C temperature increase (Lorius *et al.*, 1990). However, it is uncertain whether (1) the observed increases in CO_2 drove the warming or whether (2) warming due to planetary orbital changes drove increases in CO_2. If the first hypothesis is correct, the historic temperature sensitivity to natural CO_2 increases is comparable to that expected from anthropogenic loading of greenhouse gases (see Section III,C). If the second hypothesis is correct, current climate models may underestimate how much temperature will increase in response to anthropogenic additions of greenhouse gases to the atmosphere. If an initial warming (whether from orbital changes, greenhouse gases, or some other factor) results in more atmospheric CO_2 and other greenhouse gases, the increased gases would tend to drive additional warming. This would represent a positive feedback effect that is currently not included in climate models (see Section IV).

TABLE I

A Sample of Greenhouse Gases Affected by Human Activities[a]

	CO_2	CH_4	N_2O	CFC-11	HCFC-22
Pre-industrial concentration	~280 ppmv	~700 ppbv	~275 ppbv	0	0
1994 concentration	358 ppmv	1720 ppbv	312[b] ppbv	268[b] ppbv	110 pptv
Rate of change[f]	1.5 ppmv/yr	10 ppbv/yr	0.8 ppbv/yr	0 pptv/yr	1.2 pptv/yr
% rate of change[f]	0.4%/yr	0.6%/yr	0.25%/yr	0%/yr	5%/yr
Atmospheric lifetime (years)	50–200[c]	12[d]	120	50	12
Direct radiative forcing (Wm^{-2})	1.56	0.47	0.14	0.25[e]	

[a] Adapted from Houghton *et al.* (1996) by permission from the Secretary of the IPCC.

ppmv, parts per million by volume; ppbv, parts per billion by volume; pptv, parts per trillion by volume.

[b] Estimate from 1992–1993 data.

[c] No single lifetime for CO_2 can be defined because of different rates of uptake by different sink processes.

[d] This has been defined as an adjustment time which takes into account the indirect effect of methane on its own lifetime.

[e] This is the direct radiative forcing due to CFCs and HCFCs combined. However, their net radiative forcing is reduced by about 0.1 Wm^{-2} because they have caused stratospheric ozone depletion which gives rise to a negative radiative forcing.

[f] The growth rates of CO_2, CH_4 and N_2O are averaged over the decade beginning 1984; halocarbon growth rates are based on 1990–1994 data.

B. Recent Climate Change

Recent observational data not only document increased atmospheric concentrations of greenhouse gases, but also reveal the first "fingerprint" of human-induced global warming. Global average surface temperature trends reveal a 0.3 to 0.6°C temperature increase during the twentieth century, with the ten hottest years on instrument record occurring after 1980. Models that take into account both the positive effects of greenhouse gases and the negative effects of aerosols predict the observed upward trend of temperature quite closely (Fig. 4), including perturbations like the temporary decrease in global average temperature due to stratospheric aerosol loading from the 1991 Mt. Pinatubo volcano eruption. The current consensus among scientists is that "the balance of evidence suggests a discernible human influence on global climate" (Houghton *et al.*, 1996) (see Box 1).

How do the recent temperature increases compare to natural climate variability? The long-term climatic record is quite variable, with changes of up to 1°C per decade during the most volatile periods of transition from the last glacial to the current interglacial period. However, in the last 10,000 years of the current interglacial period, fluctuations have generally not exceeded 1°C per century, making it unlikely that the recent observed trends and expected increases in temperature over the next 100 years are merely due to natural climate variability.

C. Future Climate Change

What will the future bring? General circulation models (GCMs) integrate the physics of radiative forcing, ocean dynamics, and other complex Earth–atmosphere processes in order to simulate and predict climate trends. The current consensus, based on multiple GCM forecasts, is that a doubling of equivalent CO_2 concentration (referred to as "$2 \times CO_2$") in the atmosphere over preindustrial concentrations will result in a global mean average temperature increase of 1 to 3.5°C (Houghton *et al.*, 1996). A $2 \times CO_2$ atmosphere is used as a benchmark by climate change scientists to establish climate sensitivity and to compare scenarios and predictions, but greenhouse gas concentrations will continue to increase beyond a doubling, resulting in even greater temperature increases.

GCM and associated simulation studies point to many climatic impacts of increases in greenhouse gases and aerosols that will be of particular importance to ecosystems and biodiversity. First, surface warming will

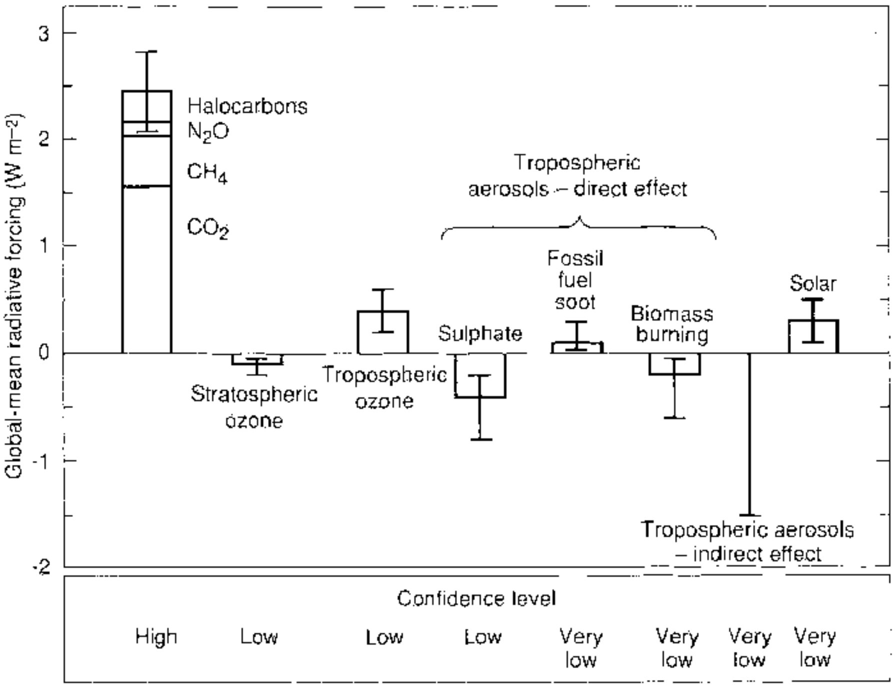


FIGURE 1 Estimates of the globally and annually averaged radiative forcing due to anthropogenic changes in concentrations of greenhouse gases and aerosols from preindustrial times to 1994 and due to natural changes in solar output from 1850 to 1994. The height of the rectangular bars indicates a midrange estimate of forcing, and the error bars show an estimate of the uncertainty range, based largely on the spread of published values. The "confidence level" indicates subjective confidence that the actual forcing lies within the error bar. (Reprinted from Schimel *et al.*, Radiative forcing of climate change, in Houghton *et al.*, 1996, by permission from the Secretary of the IPCC.)

not be uniform. Under a 2 × CO_2 atmosphere, average temperature increases are predicted to be greatest at the poles (≈10°C) and least at the equator (≤1°C). Winter temperatures will increase more than summer temperatures, and nighttime more than daytime temperatures. The incidence of record-breaking hot days will tend to increase in the summer, and fewer frost days are likely to occur in the winter. Second, warming is predicted to increase evaporation and global mean precipitation, and may increase cloud cover. High and middle latitudes and elevations are predicted to experience increases in winter precipitation, more winter precipitation falling as rain, earlier snowmelt, and reductions of summer soil moisture in noncoastal areas. Tropical precipitation is likely to change, but how it will change is uncertain. Third, the frequency and intensity of extreme weather and disturbance events (e.g., drought, deluge, summer heat waves, hurricanes, fires) are expected to increase. Fourth, glaciers are predicted to retreat and melt, sea levels may rise up to 95 cm as a result of a 2 × CO_2 atmosphere, and surface waters are likely to warm.

Some of these trends have already been observed in recent climate data. For example, more warming has occurred toward the poles, nighttime temperatures have increased more than daytime temperatures, warming has been greatest over the midlatitude continents in winter and spring, snowpack is decreasing and snowmelt is occurring earlier at high latitudes and elevations, glaciers are retreating, and sea levels have risen 10–25 cm during the last 100 years.

IV. FEEDBACKS

A. Geophysical Feedbacks

Recent general circulation models include not only mechanisms underlying direct greenhouse gas and aerosol radiative forcing, but also mechanisms underlying three large feedback processes: water vapor, snow/ice albedo, and cloud cover. Because the capacity of the atmosphere to hold water vapor increases as it warms, and because water vapor acts as a greenhouse gas to

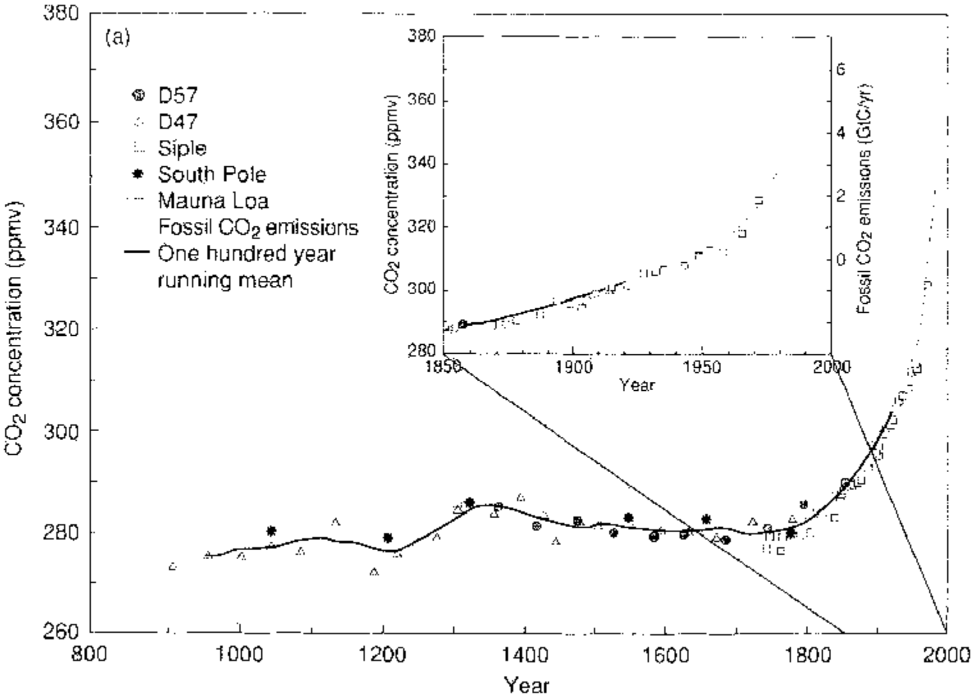


FIGURE 2 CO_2 concentrations over the past 1000 years from four Antarctic ice core records (D47, D57, Siple, and South Pole) and since 1958 from the Mauna Loa, Hawaii, measurement site. The solid curve is based on a 100-year running mean. The inset curve shows fossil fuel CO_2 emissions since 1850. (Reprinted from Houghton *et al.*, 1996, by permission from the Secretary of the IPCC.)

further increase temperature, a positive feedback to the climate is created that amplifies warming. The snow/ice albedo effect is also a positive feedback—as warmer temperatures melt highly reflective snow and ice at the poles and high elevations, thus lowering surface albedo, Earth will absorb and radiate more infrared radiation, which will augment warming. Cloud formation adds much of the uncertainty to GCM estimates because the impacts of warming and other changes to the atmosphere on cloud formation are myriad and difficult to predict. In addition, because clouds can form at many heights, over land or water surfaces with differing albedos, and with many shapes, clouds can have multiple negative and positive feedback effects.

B. Biogeochemical Feedbacks

As complex as GCMs are, they fail to incorporate many potentially important aspects of chemical and ecological processes that are likely to produce additional feedback effects. Though it is likely that none of these feedbacks is individually as strong as the geophysical feedbacks already incorporated into GCMs, together they represent a potentially significant perturbation to the climate system. Biogeochemical feedbacks could greatly increase climate sensitivity to a 2 × CO_2 atmosphere (up to 8–10°C; Lashof, 1989) compared to current model predictions (1–3.5°C). Many biogeochemical feedbacks due to global warming will also interact with other anthropogenic stresses, such as deforestation and pollution, to further exacerbate or reduce climate change effects at local, regional, and global scales.

1. Marine Feedbacks

Global warming is likely to drive many complex marine-based feedback processes, of which only a few are mentioned here. With regard to geochemistry, the warming of ocean surface waters expected under climate change will reduce the capacity of those waters to hold dissolved CO_2, resulting in a positive feedback since less CO_2 will be removed from the atmosphere. If the additional surface water warmth travels down the water column, methane could be released from temperature- and pressure-sensitive hydrates that are present in some ocean floor sediments. This would also result in a positive feedback since the released methane, a greenhouse gas, would cause more warming.

With regard to biology, more than a third of annual global primary production occurs in ocean surface waters, primarily by microscopic single-celled organisms

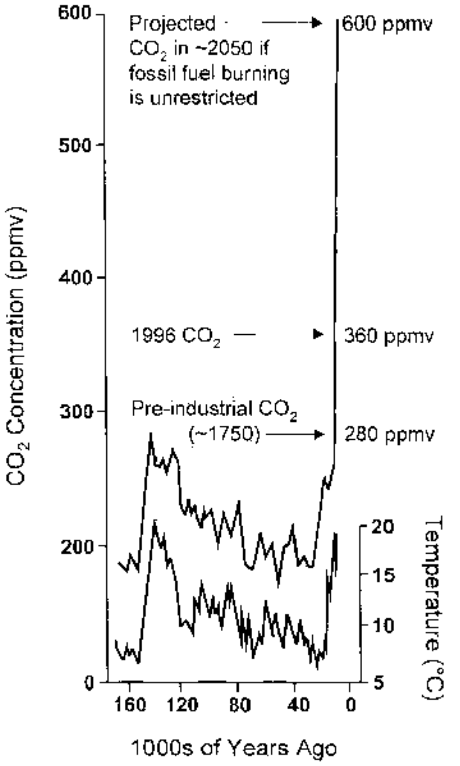


FIGURE 3 Variation during the last 160,000 years of atmospheric CO_2 concentrations and atmospheric temperature, as derived from ice core measurements. The top line shows historic, current, and projected CO_2 concentrations. The bottom line shows historic temperatures. (Adapted by permission from C. Flavin, Storm warnings: Climate change hits the insurance industry. World Watch 7, p.10. Copyright © 1994 by World Watch Institute.)

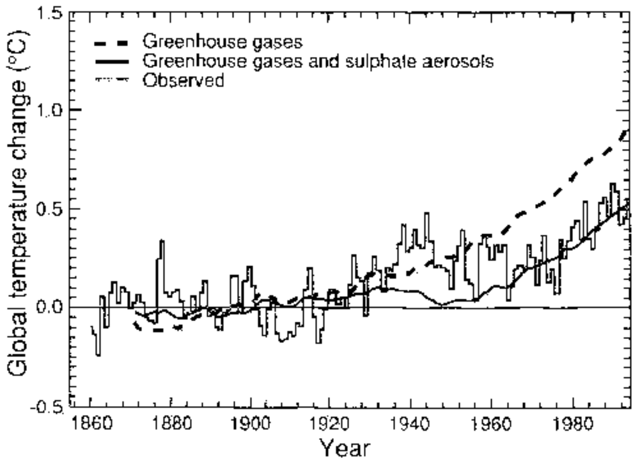


FIGURE 4 Simulated change in global mean temperature from 1860 to 1990 allowing for increases in greenhouse gases only (dashed curve) and greenhouse gases plus sulfate aerosols (solid curve), compared with observed changes over the same period. (Reprinted from Houghton *et al.*, 1996, by permission from the Secretary of the IPCC.)

Box 1

Critical Viewpoints and Responses

Although scientists currently have reached a strong consensus regarding the relationship between atmospheric loading of greenhouse gases and recent and projected global warming trends, a few critics express alternative viewpoints about climate change. Responses to some skeptical views follow:

1. *The global warming trend over the past 100 years is the result of an increasingly brighter sun.* No mechanism is known that could convert the observed very slight changes in solar output into a warming trend consistent with observations.
2. *Satellite data disprove the hypothesis that recent warming is due to greenhouse gases.* Satellite data cover too short of a period to disprove climate models, but the available satellite data (mostly on upper atmosphere temperatures) are reasonably consistent with model projections.
3. *Most of the recent warming occurred earlier in the century but the atmosphere has only dramatically changed more recently.* The warming trend since 1970 has been even more dramatic than that early in the century, consistent with the increasing rate of atmospheric greenhouse gas buildup.
4. *A cooling trend during the period 1940–1970 contradicts our climate models.* The observed cooling trend resulted from measured increasing levels of aerosols and dust in the atmosphere during that period. Although climate modelers initially failed to include atmospheric particulates in their analyses, when particles are included, the model predictions are consistent with observation.
5. *We will welcome global warming because an ice age is coming.* The timescale at which Earth will cool is over several thousand years and will not be counteracted by anthropogenic climate change. Over millennia, the current anthropogenic warming episode will slowly dissipate as humans stop loading greenhouse gases into the atmosphere and as most of the excess CO_2 is naturally sequestered in the deep oceans.

called phytoplankton, which form the base of the oceanic food web. Through the process of photosynthesis, CO_2 is fixed by the phytoplankton and thus transferred from the atmosphere to ocean surface waters. Some of that fixed carbon, in the form of dead bodies and fecal matter of phytoplankton and other organisms, sinks into deep ocean layers and sediments and is sequestered there, where it can no longer be exchanged with the atmosphere on short timescales. This process is referred to as the "biological carbon pump" and has been important in maintaining a level of CO_2 in the atmosphere that is currently about 40% lower than it would be in the absence of marine organisms. Because the amount of marine primary production is dependent on the supply of nutrients and sunlight to ocean surface layers, any way in which global warming alters those inputs can create feedbacks to warming. One of many hypothesized feedbacks, in this case positive, is that global warming will tend to diminish the intensity of the oceanic circulation of nutrients, leading to more homogeneous, diffuse ocean productivity and hence a decrease in the amount of carbon "pumped" into deep ocean layers (Rowe and Baldauf, in Woodwell and Mackenzie, 1995).

2. Terrestrial Feedbacks

About three times as much carbon is stored in terrestrial vegetation and soils than is stored in the atmosphere. Through photosynthesis and respiration, more than one-eighth of atmospheric CO_2 is exchanged each year with terrestrial ecosystems. Changes to terrestrial–atmospheric carbon cycling thus have the potential to produce significant feedbacks to climate change. The feedback pathways for carbon in terrestrial ecosystems are complex, representing both positive and negative feedbacks. Perhaps the most well-known potential carbon cycle feedback is the "CO_2 fertilization effect," which refers to the stimulation of photosynthesis by increased levels of CO_2, which in turn can result in increased plant growth and greater storage of carbon in vegetation. This represents a negative feedback to global warming. However, ecosystem-level experiments appear to indicate that fertilization effects may tend to disappear after a few years, and the magnitude of effects may be strongly impacted by water and nutrient availability (Lashof *et al.*, 1997).

Climatic changes that will accompany higher concentrations of CO_2 make it even more complicated to predict the net effect of global warming on the storage of carbon in vegetation and soil versus the atmosphere. For example, under global warming, changes in water availability and temperature will reduce the CO_2 uptake and growth of some plants while favoring others. Resulting changes in plant community composition can alter the quantity and quality of litter that enters the soil, which can lead to changes in soil carbon storage. Soil microorganisms will not only respond to changes in litter inputs, but will also be directly affected by changes in climate. Microbes and fungi tend to respire more CO_2 to the atmosphere as temperatures increase. However, rates of respiration depend on levels of soil moisture, and different extremes of water availability (both too much and too little) will tend to decrease respiration. The effects of climate change on microorganisms will also alter fluxes of other greenhouse gases such as methane in wetlands (e.g., northern peatlands, which store large amounts of carbon and may be a source of strong positive feedback to warming) and nitrous oxide in moist tropical soils.

In addition to the more direct effects of changes in temperature and moisture on the terrestrial carbon cycle, indirect effects such as the alteration of fire regimes due to climate change may produce significant feedbacks. In general, predicted increases in fire frequency for many ecosystems as a result of global warming may alter the structure of plant communities and result in reductions of terrestrial carbon storage, a positive feedback.

Another set of global warming feedbacks related to changes in albedo may result from climate-induced shifts in land cover and vegetation. The drying of soils and increased desertification expected from global warming will add dust to the atmosphere that, like aerosols, can reduce warming through increases in atmospheric albedo. Surface albedo is also expected to change as the boundaries of biomes shift, since different vegetation types can have different reflectivity. For example, the predicted northward expansion of boreal forest into tundra could decrease surface albedo, resulting in increased surface warming. This mechanism may have acted as a strong positive feedback 6000 years ago when an initial warming at high latitudes as a result of orbital variations appears to have doubled in magnitude owing to changes in surface albedo from boreal forest expansion (Foley *et al.*, 1994).

V. CLIMATE CHANGE AND BIODIVERSITY

A. Introduction

Currently, the largest reductions in biodiversity result from massive deforestation in the tropics, in conjunction with other sources of worldwide habitat destruction. Even as the razing of vast tracts of tropical forests leads to immediate direct losses of hundreds to many thousands of species per year, the carbon that is released to the atmosphere through deforestation is amplifying the anthropogenic greenhouse effect. Global warming could lead to losses in biodiversity over the next several hundred years that are similar to or greater in magnitude than losses from direct habitat destruction. Dramatic changes in global climate have the potential to disrupt every ecosystem on Earth, leading to a pervasive trend of biodiversity loss due to climate-related habitat alteration, reorganization, and destruction.

Why does anthropogenic climate change present such a threat to biodiversity? There are two main reasons. First, the rate and magnitude of climate change expected over the next several decades to centuries are greater than any changes that current organisms have experienced. Over the last 18,000 years, starting during the last full glacial period and continuing through the current interglacial period that began about 10,000 years ago, average global surface temperature has gradually increased by about 5 ± 1°C (T. Webb, in Peters and Lovejoy, 1992). If we assume conservatively that global warming will increase mean temperature by 5°C over the next 200 years, this represents a 90-fold increase over the recent natural rate of change. In terms of magnitude, a 3°C increase would result in the warmest world in 100,000 years, and a 4°C increase would result in the warmest world in 40 million or more years. Second, global warming will interact synergistically with other anthropogenic stresses such as habitat destruction, pollution, ozone depletion, and alien species introduction to reduce biodiversity by more than just the sum of losses that would occur if each factor occurred independently.

Scientists' ability to specifically predict how biodiversity will be affected by climate change is constrained by large uncertainties associated with local and regional climate change predictions. Much of the following discussion is based not so much on how scientists specifically think biodiversity will change under global warming, but on generic ways in which biota will be affected by global warming that can lead to changes in biodiversity. The term "biota" is used as shorthand to refer collectively to individual organisms, groups of the same types of organisms (populations, species, functional types), and ecological complexes of multiple populations and species (communities, ecosystems). This section summarizes how scientists study the effects of climate change on biota, the types of responses biota can have to climate change, the kinds of biota likely to be harmed by and to benefit from climate change, and evidence for biotic responses to current anthropogenic climate change.

B. How Scientists Study the Effects of Climate Change on Biota

1. Models

Climate–vegetation classification systems are types of static models that are based on the hypothesis that climate patterns are the primary determinant of the broad-scale distribution of vegetation types. In 1947, L. Holdridge developed a "life-zone" concept that used three variables based on temperature and precipitation to predict under what climates 20 vegetation types should occur. Later researchers refined and added detail to this basic concept by using a wider variety of vegetation types and bioclimatic variables that explicitly incorporate drought stress and seasonality (Table II). These types of static models can be compared to current vegetation and climate maps to determine their accuracy and then used in conjunction with maps of simulated future climate to predict shifts in distribution of vegetation types due to global warming. The Holdridge life-zone classification has been used to predict the conversion of much of today's boreal forest into temperate deciduous forest (Emanuel *et al.*, 1985). While classification systems provide ways to look at potential global-scale impacts of climate change on vegetation, they are limited by the fact that climate will change continuously and with considerable interannual variation, rather than shifting abruptly to a new plateau. In addition, many other factors besides climate can influence the movement and distribution of biota, and this greatly reduces the potential accuracy of such models.

TABLE II

Bioclimatic Variables Used by E. O. Box to Predict Distribution Limits of Plant Types[a]

T_{max}	Mean temperature of the warmest month (°C)
T_{min}	Mean temperature of the coldest month (°C)
D_T	Range between T_{max} and T_{min} (°C)
P	Mean total annual precipitation (mm)
P_{max}	Mean total precipitation of the wettest month (mm)
P_{min}	Mean total precipitation of the driest month (mm)
P_{Tmax}	Mean total precipitation of the warmest month (mm)
MI	Moisture index: the ratio of P to annual potential evapotranspiration

[a] From 1981, Macroclimate and Plant Form, Junk Publishers, The Hague.

Simulation models incorporate and link multiple factors such as life-history traits, physiological constraints, biotic interactions, resource availability, and climate in order to predict dynamic changes in biota over the course of time. By using characteristics of individuals, populations, species, or functional types in conjunction with predicted changes in climate, modelers can simulate continuous vegetation responses to global warming at scales that vary from changes within a stand of plants to changes at landscape, regional, and global levels. Vegetation models can be used in turn to look at changes in habitat and dynamics of other types of organisms. Simulation models not only attempt to predict patterns of change, but also provide a means to explore how and why such change might occur. However, these analyses are often constrained by data requirements. Detailed and accurate information may not be available about the biota and environment of interest, and as simulations are scaled up to look at global impacts, specificity is lost. No matter how well models appear to fit observed dynamics, some factor not included in the model may be important for future dynamics, or the model may fit observed dynamics for the wrong reasons, rendering model predictions inaccurate.

2. Paleobiology

Paleobiologists combine historical climate and biological data to reconstruct past relationships between changes in climate and species distributions. Paleobiological data sources range from ice and soil cores to plant and animal fossils to tree rings and slow-growing corals. Such studies reveal the long-term and integrated direct and indirect effects of past climatic and atmospheric change on past biota and thus facilitate projections about how present-day biota might respond in the long term to anthropogenic climate change. Many crucial insights about previous and potential effects of climate change have emerged from this type of work (see Section V,C). However, the generality of these types of studies is limited by the lack of good fossil records for many organisms, by the lack of strict climatic parallels in the past to both the rate and magnitude of anthropogenic climate change, and by differences in the biology and geology of ancient times compared to the present. In addition, the interactions of global warming with other anthropogenic stresses such as pollution, development, agriculture, and deforestation are novel and not represented in the paleorecord.

3. Natural Climate Variability

Whereas paleobiologists look to the ancient past for insight into the future, many field biologists interested in global warming look to the present to examine how biota are regulated by and respond to natural climate variability of a magnitude similar to that expected from global warming. Two approaches are used in this type of research. First, scientists can conduct "space-for-time" analyses along elevational or latitudinal gradients. This approach suggests that the effects of climate change over time on particular biota may be represented by current differences between the biota of interest and the same type of biota found at warmer, lower elevations or latitudes. Second, researchers who conduct multiyear studies at particular sites can monitor the response of biota to the natural interannual variability of climate. In particular, biotic responses to climate in more "normal" years can be compared to biotic responses to very warm years, droughts, early snowmelts, and other climatic events that fit predicted changes due to global warming. In some cases, researchers have access to records of climate and biota from several decades ago and can compare them to current records for the same sites.

This type of research has the advantage of actually studying current biota in the field in relation to climate, but it also has several disadvantages. For example, from one year to the next and from one place to another, sites vary not only in their climate but also in many other factors (e.g., land use history, species composition, topography), making it difficult to establish whether particular climate factors and/or other nonclimatic factors underlie observed patterns. Even if sites appear to differ primarily in climate, they may differ by predicted amounts of temperature but not by predicted changes in precipitation and soil moisture. Additionally, biota have had longer time periods to adjust and adapt to current climate variation at particular sites than they will have to adjust to climate change from global warming. Finally, this approach is unable to explore the effects of increased atmospheric CO_2, with the exception of some research on natural CO_2 gradients near hot springs.

4. Manipulations

Manipulations are one of scientists' most potent research tools. By conducting controlled manipulations of various climate, atmosphere, and other resource factors expected to change as a result of global warming, climate change researchers can work toward understanding the role of single or multiple factors and their interactions in changing ecosystem structure and function. This type of research is used to predict both how specific anthropogenic climate change scenarios might impact biota and how resulting ecosystem changes may produce feedbacks to the climate system. In climate change ecology, manipulative research falls into two types of approaches: microcosm experiments and field experiments. Microcosms, which generally take the form of laboratory growth chambers of various sizes, can be used to carefully manipulate particular global warming factors (e.g., temperature, moisture, light, nutrients, atmospheric composition) and to monitor the response of soils, single or multiple organisms, assembled simple ecosystems, or intact ecosystem cores taken from the field. Field experiments also manipulate global warming factors in order to look at interactions between climate change and biota, but do so in intact natural ecosystems (Fig. 5).

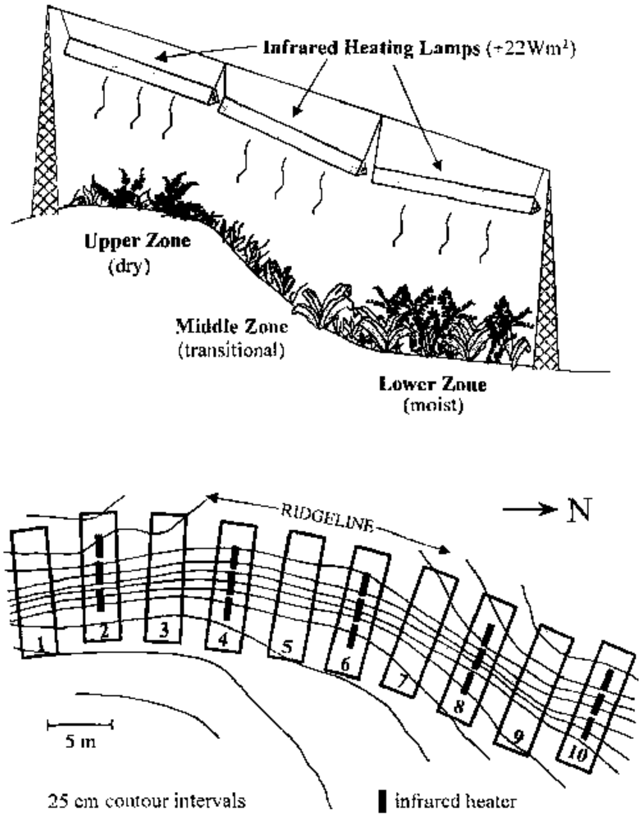


FIGURE 5 Profile of a heated plot used in a long-term ecosystem warming experiment to investigate the interactions between climate warming and the ecology and biogeochemistry of a subalpine meadow ecosystem at the Rocky Mountain Biological Laboratory in Gothic, Colorado, U.S.A. The heaters increase downward infrared radiation flux (+22 W m^{-2}) to a degree comparable to that expected under a 2 × CO_2 scenario. Layout of the plots is shown below; even-numbered plots have heaters and odd-numbered plots are control plots that lack heaters. (Adapted with permission from J. Harte and R. Shaw, Shifting dominance within a montane vegetation community: Results of a climate-warming experiment. *Science* 267, 876. Copyright © 1995 by AAAS.)

The strengths and weaknesses of these two approaches are interrelated. Although it is relatively easy to manipulate, control, and replicate microcosms, field experiments may be confounded by ecosystem variability and complexity. In field experiments, typically only a very few experimental variables can be manipulated, controls can be difficult to establish, and adequate replication is often expensive and time-consuming. However, field experiments have the advantage of being conducted in a natural setting at broader scales that may be more useful for drawing conclusions about complex, "real-world" ecosystem dynamics, compared to highly simplified and small-scale microcosm experiments. In both types of experiments, the relatively abrupt, short-term manipulation of climate may not be a good analog of anthropogenic climate change, which is occurring more gradually over decades and centuries. In addition, changes to disturbances such as fires and hurricanes may prove to be more important in determining the abundance and distribution of biota in many ecosystems than the usual experimental focus on "average" changes in variables such as temperature, moisture, and CO_2.

5. Integrated Research

Given the limitations of each type of research, the most productive strategies for exploring interactions between ecosystems and anthropogenic climate change integrate multiple research approaches. For example, responses of biota to natural climate variation can be compared to responses of the same biota to manipulated climate change to see how responses differ or stay the same over multiple spatial and temporal scales. Results from field experiments and gradient studies often suggest mechanisms that can be more thoroughly tested in microcosm experiments. Gradient, field experiment, and paleobiological data sets can be used to parameterize, calibrate, and validate models of biotic response. In general, the thoughtful integration of approaches can build on the strengths and avoid some of the limitations of each type of research, thus helping scientists to develop more rigorous hypotheses about global warming impacts and ecosystem feedbacks.

C. Impacts of Climate Change on Biota

1. Types of Responses of Organisms, Populations, and Species

a. Adjustment

The first level, short-term response of any organism to changes in their environment is adjustment, also referred to as acclimatization. All organisms have some degree of physiological, life-history, or behavioral plasticity that enables them to live in a variable environment. The degree of plasticity with regard to climatic and atmospheric conditions varies widely among different kinds of organisms. Therefore, some types of organisms will be able to adjust to relatively large changes in climate, whereas others will be unable to adjust to even apparently minor increases in temperature or slight variations in precipitation.

An example of climatic adjustment in animals involves thermoregulation in vertebrates. Endotherms such as mammals have built-in physiological mechanisms to cope with body temperature changes. Ectotherms such as reptiles have behavioral traits that help regulate body temperature. Because of traits such as these, initial increases in environmental temperature should be well within the tolerances of many vertebrates. In plants, the concurrent increase of atmospheric CO_2 with surface temperature may augment the ability of some individuals, populations, and species to adjust to and flourish under anthropogenic climate change. Increases in CO_2, especially for plants with the common C_3 photosynthesis pathway (e.g., most trees and shrubs), can result, at least initially, in the CO_2 fertilization effect mentioned before. Enhanced CO_2 concentrations can increase the ability of these types of plants to tolerate water stress, higher temperatures, and lower light. Other kinds of plants, particularly those with the C_4 photosynthesis pathway (e.g., many low-latitude and low-elevation grasses), have physiological mechanisms that enable them to withstand warm temperatures and low availability of water. Such mechanisms provide a means of adjustment to drought stress that may be associated with increased temperatures and evaporation.

While most biota will have at least some capacity to withstand, and in some cases benefit from, initial changes in climate, the rapid rate and large magnitude of climate change are likely to quickly surpass their capacity to adjust to new climate conditions within their pre-warming habitats. Biota that cannot continue to adjust will have to respond through evolution, migration, or extinction.

b. Evolution

Theoretically, populations and species could develop new adaptive traits as a result of evolution in response to anthropogenic climate change, thus enhancing the long-term survival of current taxa under new climate conditions. However, scientists generally agree that evolutionary responses to climate change are unlikely for most taxa since climate is changing rapidly compared to usual rates of evolutionary change. This view is supported by fossil data that reveal the morphological stasis of many taxa during previous periods of rapid climate change.

In the face of strong selectional pressure there is evidence that some species, especially those with fast generation times, can evolve very rapidly. For example, grass populations grown on soils polluted by heavy metals have shown signs of significant, genetic-based, heavy-metal tolerance within one or two decades. Some insects can evolve increased resistance to pesticides over the course of a few years. These types of responses depend on the presence of appropriate genetic variability in populations and species relative to a strong selective factor. As a result of global warming, populations and species will be exposed to novel environments resulting from climate change and associated shifts in ecosystem structure and function. Since many populations and species have climate-related genetic variability (e.g., differences in high temperature tolerance, drought tolerance), rapid adaptation is possible.

Will anthropogenic climate change actually result in directed selectional pressures that are strong enough to drive microevolutionary responses? For at least the next several hundred years, climate and species distributions are likely to be in fairly constant flux, which will tend to disrupt any potentially adaptive trends. Also, rapid evolutionary responses to anthropogenic climate change are unlikely in populations and species that have relatively long generation times, such as trees and many vertebrates. An added constraint on potential microevolutionary responses to global warming is the ongoing reduction in population size and thus genetic diversity of many species as a result of habitat destruction and other stresses. For most biota, other types of responses are far more likely to occur than evolution.

c. Migration

As current habitat becomes inhospitable owing to direct and indirect effects of climate change, biota will tend to track shifting climate and suitable habitat through dispersal and migration. Consequently, as a result of global warming, organisms are predicted to move generally poleward in latitude and upward in elevation. A rule of thumb is that a 3°C change in temperature is approximately equivalent to a move of 250 km of latitude or 500 m of elevation. However, migration will be restricted or made impossible to the extent that there are inherent (e.g., low mobility, slow reproductive rates) or external barriers (e.g., mountain ranges, large lakes) to movement (see Boxes 2 and 3).

Box 2

Biota Most at Risk from Climate Change

Given our knowledge of climate change and biology, the kinds of biota most likely to be at risk from global warming over the next several decades and centuries can be characterized:

1. *Those at higher latitudes:* Scientists know with a high degree of certainty that temperature increases due to global warming will be greatest in polar regions. Therefore, higher-latitude temperate and Arctic/Antarctic ecosystems such as boreal forest, tundra, and peat bogs will experience both rapid and severe temperature increases, resulting in profound biotic change and disruption.
2. *Those on mountain-tops:* Temperature increases will also be greater at higher elevations. Montane biota will tend to move up in elevation as cooler, higher elevations warm. Biota already limited to mountaintops will be at serious risk of local extinction due to alteration of summit climate, the lack of potentially suitable habitat to migrate to, and the encroachment of lower-elevation species.
3. *Those in low-lying coastal areas and on islands:* Even small increases in sea level (i.e., several centimeters) can result in altered coastal marine dynamics and flooding of low-lying areas. Rising sea levels will destroy or cause severe damage to ecosystems at the terrestrial/marine interface, such as salt marshes, estuaries, mangroves, and sand dunes, and are likely to disrupt coastal marine food webs.
4. *Those sensitive to extreme disturbance events:* Even though disturbances such as fires and hurricanes are a natural part of ecosystem dynamics, any increases in frequency and intensity of such disturbances due to global warming are likely to disrupt biota. For example, ecosystems such as tropical montane forests may have less time to recover between hurricanes, limiting the development of slow-growing, late-successional species.
5. *Those with migration problems:* Because a major potential response of biota to climate change is to migrate to new areas, biota that lack the ability to readily disperse or move will be at a serious disadvantage. These include:
    - (a) plants whose seed or clone dispersal rates and animals whose movement rates lag behind rates of climate change;
    - (b) slow-growing populations and species that will not have time to adjust to new conditions;
    - (c) biota that cannot or are slow to cross geographic barriers, for example, fish in isolated lakes, low-elevation plants bounded to the north by mountain ranges, and tropical forest birds and insects that do not cross unforested areas;
    - (d) organisms that depend on other biota for habitat or food, but that have very different degrees of mobility;
    - (e) species such as monarch butterflies and migratory shorebirds that have multiple habitat requirements; and
    - (f) relic biota that have been left in small, unusual habitats by chance and have no nearby potential habitat.
6. *Those that are rare:* Populations and species with few numbers, low genetic variability, or limited or unusual ranges will be vulnerable to climate fluctuations and will be at increased risk of extinction. Also, unusual, unique ecosystems may break apart as populations and species respond in largely individualistic ways to climate change.
7. *Those dependent on particular hydrological regimes:* Though it is often uncertain at local and regional scales how and to what degree precipitation and moisture availability will change from global warming, it is quite certain that change will occur. Such changes could be critical in ecosystems such as tropical forests where the availability of food resources for animals is dependent on the timing of rainfall. In montane areas, many organisms will be very sensitive to changes to the snowpack and snowmelt.
8. *Those close to critical physiological thresholds:* Many organisms are adapted to living within a narrow range of limits of temperature, moisture, nutrients, light, and atmospheric composition. Others have wider tolerances but are already operating close to a threshold, beyond which their ability to live, grow, and reproduce is severely limited. Climate change may force the environment past such limits for some organisms, resulting in severe im-

pacts, especially if dispersal or growth is slow. For example, slight increases (1–2°C) in surface water temperature can induce bleaching and mortality of coral reefs, which have very slow rates of growth and provide habitat for many marine species.

9. *Those that have highly specialized relationships with other organisms:* Some species depend entirely on just one or a very few other species for nourishment or reproduction. If species respond very differently to changes in climate than do the species they depend on, and they cannot substitute other organisms to fulfill those roles, they will go extinct.
10. *Those negatively affected by other anthropogenic stresses:* Humans engage in many activities that result in deleterious impacts such as acid deposition, pollution, ozone depletion, and alien species introduction. When organisms are weakened by one of these stresses, they tend to become even more vulnerable to other stresses such as global warming. For example, insect pests can damage vegetation more when pollutants reduce plants' resistance to herbivory and warmer temperatures encourage pest population growth. If those insect pests happen to be alien, they may cause even more damage owing to lack of local predators and because local plants may lack resistance to alien pests. Also, changes in land use such as deforestation can reduce and isolate populations as well as create barriers to migration through habitat fragmentation and destruction.

Box 3

Biota That May Benefit from Climate Change

Because of the slow rate of most evolutionary responses, including speciation, there are few ways that global warming could augment global biodiversity over the next several hundred years. However, climate change will benefit some biota through increases in abundance and range expansions, often at the expense of more at-risk biota. Types of biota likely to benefit from climate change include:

1. *Those that migrate easily:* Biota that are highly mobile and have rapid dispersal rates, such as some kinds of winged insects, will be equipped to track changing climate.
2. *Those that are opportunistic:* Opportunistic organisms that can colonize disturbed areas will be at an advantage because they will be able to migrate through marginal habitat and establish in climatically disrupted ecosystems. For example, global warming is expected to promote the spread of already weedy introduced plant species, and may facilitate the escape of more garden cultivars into natural ecosystems.
3. *Those that are ecological generalists:* Organisms that flourish in a wide variety of environments and have either wide tolerances for variable resource availability and climate or many possible prey items will fare better than highly specialized organisms.
4. *Those that have high variability and rapid reproduction:* Populations and species with lots of phenotypic or genetic variation and rapid reproductive rates have the best chances of adjusting and adapting to rapidly changing climate.
5. *Those favored by new optima:* Although climate and atmospheric conditions will shift away from optima for many organisms, more optimal conditions will be created for other organisms. For example, temperature increases are expected to increase parasite and insect development time, allowing parasites to spread with migrating insect hosts and promoting pest infestation and parasite infection of new hosts. These types of responses may lead to range expansions of agricultural pests and disease transmissions, as well as more frequent outbreaks (Dobson and Carper, in Peters and Lovejoy, 1992).

Fossil records show that for many types of organisms, warming during the last deglaciation induced significant changes in latitude and elevation of species' ranges (Fig. 6). Those distributional changes sometimes occurred at very rapid rates. For example, peak migration rates for some tree species in North America during the last deglaciation reached 100–500 m per year, probably as a result of haphazard, long-distance transport of seeds by animals, storms, or water (Clark *et al.*, 1998). However, even these very fast historic migration rates only translate into 10–50 km per century, whereas anthropogenic climate change will likely require latitudinal shifts of at least 200–300 km over the next century. In some cases, changes in potential range bound-

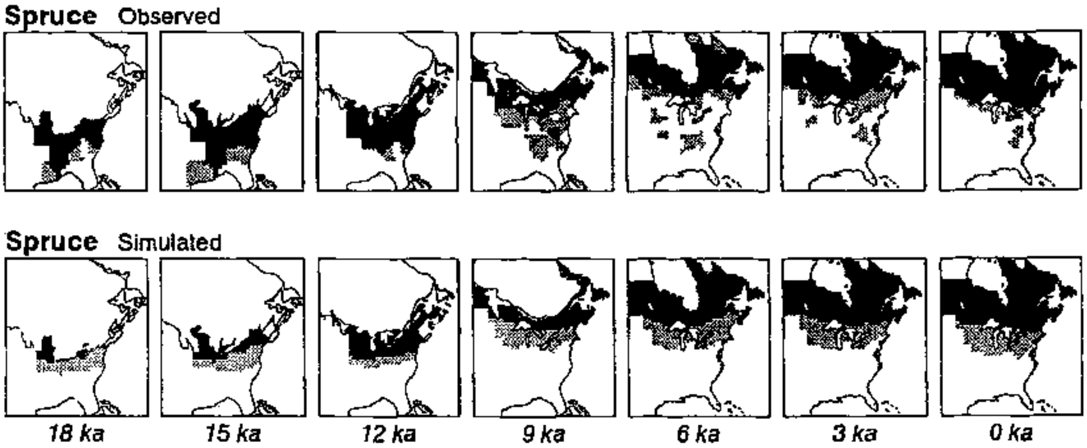


FIGURE 6 Maps of observed (upper row) and simulated (lower row) percentages for spruce tree pollen in eastern North America over the last 18,000 years. The simulated spruce pollen maps are based on the modern response of spruce pollen percentages to July temperature, January temperature, and precipitation as applied to simulated historical climates. Dark shading indicates the highest abundance of spruce. (Reprinted with permission from T. Webb, III. Past changes in vegetation and climate: Lessons for the future, in Peters and Lovejoy, 1992. As adapted with permission from COHMAP Members. (1988). Climatic changes of the last 18,000 years: Observations and model simulations. *Science* 241, 1043. Copyright © 1988 by AAAS.)

aries resulting from climate change due to doubled CO_2 may exceed 1000 km. For example, suitable habitat for beech trees, which currently grow over most of the eastern third of the United States, could shift almost completely out of the country and into a much smaller area of the northeastern United States and Southeastern Canada as a result of a 2 × CO_2 atmosphere (Fig. 7).

The rapid pace of anthropogenic climate change will easily outstrip the capacity of some organisms to move or disperse to suitable new habitat. In addition, human destruction of habitat will create insurmountable barriers to migration for biota that have difficulty crossing large areas of urban development or agricultural use. In the short run, any migration that does occur may tend to increase local levels of biodiversity in some ecosystems as new species move in before old species completely migrate or die out. This lag effect will tend to disappear with time.

d. Extinction

Biota that are unable to adjust, evolve, or migrate are unlikely to survive long in their pre-warming habitats. As a result of climate change, organisms may be exposed to increasing physiological stress, be abandoned by their mutualists and prey, be outcompeted by more flexible neighbors or incoming species, and be attacked by new predators and pests. Although microevolutionary responses are possible for some populations, rapid genetic adaptation is likely to occur in very few. Many species will be faced with migration problems. Even in species that successfully migrate, some populations are likely to go extinct, particularly at southern and lower edges of species' ranges, reducing genetic variability. Extinctions attributable primarily to global warming will be few over the next several decades, but will undoubtedly increase dramatically as time passes, climate change intensifies, and biotic response options narrow.

2. Impacts on Biotic Assemblages

A key insight from the fossil record is that species tend to respond to climate change individualistically rather than as a group. Thus, while the general trend is for species to move poleward and to higher elevations as climate warms, particular species can vary quite dramatically in how fast and how much their ranges contract, expand, or move, what directions they move in, and at what rate they move around or over barriers such as mountain ranges. Consequently, past communities repeatedly disassociated and re-sorted into novel combinations. In some cases, although the same species still exist, there are no modern examples of historic species associations. For example, for several thousand years at the end of the last glacial period, spruce trees grew in open parklands in association with sedges. Today spruce is found in a completely different ecosystem type, the closed-canopy boreal forest, in association with birch, alder, and fir (Webb, in Peters and Lovejoy, 1992). Anthropogenic climate change is likely to result in the reconstitution of communities and ecosystems in unexpected ways.

Biota tend to move individualistically in response to long-term climate change because tolerances to climatic and atmospheric conditions are often specific to the organism, population, or species. Individualistic responses can lead to apparently counterintuitive shifts in range, especially if temperature is not the primary determinant of distribution. For example, the distribution of the gopher tortoise during the most recent deglaciation shifted south, rather than north. One hypothesis for this pattern is that seasonal climate extremes increased with warming, and that these extremes were more important for determining tortoise distribution than warming (Graham, in Peters and Lovejoy, 1992). In addition to climate and atmosphere, other abiotic factors (e.g., soil type, topography, disturbance regime, site history) and biotic interactions (e.g., mutualism, competition, predation, pollination) can play significant roles in determining the abundance and distribution of biota in both the short and long term. The interplay of all of these factors over time can result in complex and often unpredictable changes in the distribution, abundance, and diversity of biota.

Global warming will also precipitate many asynchronies that reduce the ability of biota to respond effectively to climate change. "Asynchrony" refers to a mismatch in timing or rate of change. One type of asynchrony already discussed is the mismatch between very rapid rates of anthropogenic climate change and slower rates of dispersal and migration for many species. Another type of asynchrony due to global warming is the potential mismatch between required resources and the availability of resources. Organisms are embedded within a network of relationships with other organisms, which they may depend on for sustenance and reproduction. However, different types of organisms may be affected very differently by climate change, which can disrupt biotic relationships that are important for community and ecosystem dynamics. For example, many species of flowering plants depend on specific animal pollinators such as butterflies, hummingbirds, and bees for successful reproduction, and the pollinators depend on those plants for food. In a hypothetical example, if the timing of flowering in a plant is primarily determined by early season temperature, while the emergence and activity of its pollinator is primarily determined by the amount of daylight, the shifting of temperature but not light levels due to global warming could lead to the pollinators being active after the plants flower. This type of

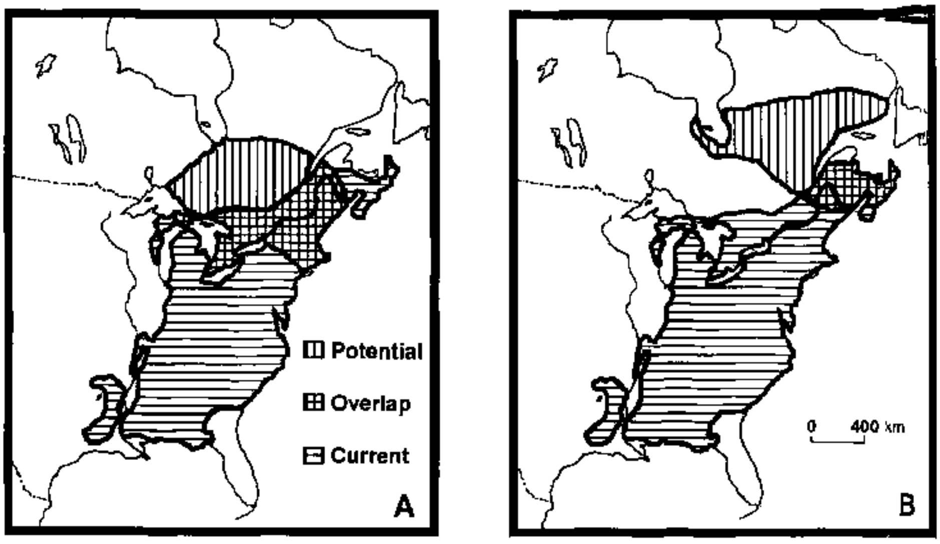


FIGURE 7 Current geographical range (horizontal lines) and potentially suitable range under doubled CO_2 (vertical lines) for beech trees in North America. Cross-hatched lines indicate areas of overlap between current and potential future ranges. (A) Output for a milder climate change scenario and (B) output for a more severe climate change scenario. (Adapted with permission from M. B. Davis and C. Zabinski, 1992, Changes in geographical range resulting from greenhouse warming: Effects on biodiversity in forests, in Peters and Lovejoy, 1992. Copyright © 1992 by Yale University Press.)

asynchrony can reduce the reproduction and abundance of the plants, and reduce the availability of food for the pollinators. If the species involved are highly specialized on each other, they are likely to go extinct.

D. Evidence for Current Global Warming Impacts on Biota

During the 1990s, researchers started reporting the first evidence that global warming during the twentieth century has begun to influence populations, species, and ecosystems (see Box 4 for a bibliography of the sources cited in this section). Some evidence comes from coastal marine systems. Roemmich and McGowan (1995) reported that since 1951, zooplankton biomass in coastal southern California waters had decreased by 80% over four decades, at the same time that surface water layers warmed more than 1.5°C in some areas. They suggested that the surface warming resulted in changes to stratification and the thermocline that led to a reduction in upwelling of nutrients and thus primary production by phytoplankton, the ultimate food source of zooplankton. Barry and colleagues (1995) compared changes in abundances of 45 invertebrate species in a central California intertidal community from the 1930s to the 1990s. During that period of time, annual mean shoreline temperature at the study area increased by 0.75°C, and average summer temperature increased by 2.2°C. Eight of nine species with a southern geographic range showed significant increases in abundance, whereas five of eight species with a northern distribution showed significant decreases. Species with wide distributions showed no strong patterns of change.

Box 4

Selected Bibliography for Current Global Warming Impacts on Biota

Barry, J. P., Baxter, C. H., Sagarin, R. D., and Gilman, S. E. (1995). Climate-related, long-term faunal changes in a California rocky intertidal community. *Science* 267, 672.

Brown, J. L., Li, S. H., and Bhagabati, N. (1999). Long-term trend toward earlier breeding in an American bird: A response to global warming? Proc. Nat. Acad. Sci. *U.S.A.* 96, 5565.

Crick, H. Q. P., and Sparks, T. H. (1999). Climate change related to egg-laying trends. *Nature* 399, 423.

Grabherr, G., Gottfried, M., and Pauli, H. (1994). Climate effects on mountain plants. *Nature* 369, 448.

Myneni, R. B., Keeling, C. D., Tucker, C. J., Asrar, G., and Nemani, R. R. (1997). Increased plant growth in the northern high latitudes from 1981 to 1991. *Nature* 386, 698.

Parmesan, C. (1996). Climate and species' range. *Nature* 382, 765.

Parmesan, C., Ryrholm, N., Stefanescu, S., Hill, J. K., Thomas, C. D., Descimon, H., Huntley, B., Kaila, L., Kullberg, J., Tammaru, T., Tennent, W. J., Thomas, J. A., and Warren, M. (1999). Poleward shifts in geographical ranges of butterfly species associated with regional warming. *Nature* 399, 579.

Roemmich, D., and McGowan, J. (1995). Climatic warming and the decline of zooplankton in the California current. *Science* 267, 1324.

Thomas, C. D., and Lennon, J. J. (1999). Birds extend their range northwards. *Nature* 399, 213.

At high elevations and latitudes, terrestrial vegetation appears to be responding to the warmer climate. Grabherr and colleagues (1994) compared 40- to 90-year-old records of plant distribution on high-altitude mountaintops (~2900 to 3500 m) in the Alps to 1992 data, given a 0.7°C increase in mean annual temperature during that time period. They found upward migration rates by nine "typical" species of <1 to 4 m per decade. In addition, species richness increased over time, with the increase most pronounced at lower-altitude summits. Myneni and colleagues (1997) presented evidence from satellite data from 1981 to 1991 that suggests that the photosynthetic activity of terrestrial vegetation in northern high latitudes increased by 7–14% over that time period. This increase may indicate an intensification of plant growth associated with a 12-day increase in the growing season.

Insect species are also displaying sensitivity to recent climate change. Parmesan (1996) examined historical records and recorded the current status of 151 populations of Edith's checkerspot butterfly throughout its entire range. She found that net extinctions of populations were significantly greater at southern latitudes and lower altitudes, resulting in an observed northward and upward shift in the species' range during the last several decades. In a later study that examined range shifts in 35 nonmigratory European butterfly species during the last century, Parmesan and colleagues (1999) reported that 63% of the species displayed northward range shifts of 35 to 240 km, while only 3% shifted south.

Vertebrates also appear to be sensitive to recent climate change. Current evidence suggests that birds are both breeding earlier and shifting their ranges northward in response to warming. Brown and colleagues (1999) found that from 1971 to 1998, the average timing of first clutch in the Mexican jay in Arizona occurred 10 days earlier. This change was associated with significant increases in monthly minimum temperatures. Crick and Sparks (1999) studied 20 United Kingdom bird species over 25 years and found long-term trends toward earlier egg laying in most species. A data set of 36 bird species over 57 years suggests that 86% of the species display significant relationships between timing of egg laying and temperature or rainfall. Furthermore, Thomas and Lennon (1999) examined the breeding distributions of British birds over a recent 20-year period and found that the northern margins of the species' ranges moved north by an average of 19 km.

In general, the observations of all of these studies are consistent with predictions of impacts of climate change on biota, and demonstrate how just a small amount of climate change over brief time periods can lead to significant ecological changes.

VI. CONCLUSIONS

The enhanced greenhouse effect, in conjunction with other anthropogenic stresses, is likely to precipitate unprecedented changes to Earth's climate and ecosystems. Though the details of how climate change will affect biodiversity are often hard to predict, there is little doubt that biological impacts will be pervasive and often dramatic. Studying the effects of climate change on biota can help in the formulation of strategies for conserving biodiversity and ecosystem structure and function in the face of potentially massive change and loss. Such knowledge is also crucial for refining predictions of the future rate and magnitude of global warming, since biological responses are likely to produce significant feedbacks that can augment or dampen climate change at local, regional, and global scales. Understanding and addressing the interactions between climate change and biodiversity represents one of the greatest challenges that scientists and policymakers will face in the twenty-first century.

See Also the Following Articles

ATMOSPHERIC GASES • BIOGEOCHEMICAL CYCLES • CARBON CYCLE • CLIMATE CHANGE AND ECOLOGY, SYNERGISM OF • CLIMATE, EFFECTS OF • DEFORESTATION • MIGRATION • SOIL BIOTA, SYSTEMS AND PROCESSES

Bibliography

Clark, J. S., Fastie, C., Hurtt, G., Jackson, S. T., Johnson, C., King, G. A., Lewis, M., Lynch, J., Pacala, S., Prentice, C., Schupp, E. W., Webb, T., III, and Wyckoff, P. (1998). Reid's paradox of rapid plant migration. *Bioscience* 48, 13.

Emanuel, W. R., Shugart, H. H., and Stevenson, M. P. (1985). Climatic change and the broad-scale distribution of terrestrial ecosystem complexes. *Climatic Change* 7, 29.

Foley, J. A., Kutzbach, J. E., Coe, M. T., and Levins, S. (1994). Feedbacks between climate and boreal forests during the mid-Holocene. *Nature* 371, 52.

Houghton, J. T., Meira Filho, L. G., Callander, B. A., Harris, N., Kattenberg, A., and Maskell, K. (eds.). (1996). *Climate Change 1995—The Science of Climate Change*. Contribution of Working Group I to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom.

Kareiva, P., Kingsolver, J., and Huey, R. (eds.). (1993). *Biotic Interactions and Global Change*. Sinauer Associates, Sunderland, Massachusetts.

Lashof, D. A. (1989). The dynamic greenhouse: Feedback processes that may influence future concentrations of atmospheric trace gases and climatic change. *Climatic Change* 14, 213.

Lashof, D. A., DeAngelo, B. J., Saleska, S. R., and Harte, J. (1997). Terrestrial ecosystem feedbacks to global climate change. *Annu. Rev. Energy and the Environment* 22, 75.

Lorius, C., Jouzel, J., Raynaud, D., Hansen, J., and Le Treut, H. (1990). The ice-core record: Climate sensitivity and future greenhouse warming. *Nature* 347, 139.

Markham, A. (1996). Potential impacts of climate change on ecosystems: A review of implications for policy makers and conservation biologists. *Climate Research* 6, 179.

Markham, A., Dudley, N., and Stolton, S. (1993). *Some Like It Hot: Climate Change, Biodiversity and the Survival of Species*. World Wildlife Fund International, Gland, Switzerland.

Peters, R. L., and Lovejoy, T. E. (eds.). (1992). *Global Warming and Biological Diversity*. Yale University Press, New Haven, Connecticut.

Schlesinger, W. H. (1991). *Biogeochemistry: An Analysis of Global Change*. Academic Press, San Diego, California.

Solomon, A. M., and Shugart, H. H. (1993). *Vegetation Dynamics and Global Change*. Chapman & Hall, New York.

Watson, R. T., Zinyowera, M. C., and Moss, R. H. (eds.). (1996). *Climate Change 1995—Impacts, Adaptations and Mitigation of Climate Change: Scientific-Technical Analyses*. Contribution of Working Group II to the Second Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge University Press, Cambridge, United Kingdom.

Woodwell, G. M., and Mackenzie, F. T. (eds.). (1995). *Biotic Feedbacks in the Global Climate System: Will the Warming Feed the Warming?* Oxford University Press, New York.

GUILDS

Richard B. Root
Cornell University

GLOSSARY

adaptive syndrome A coordinated set of adaptations.

community An assemblage of populations that coexist in an area.

Eltonian niche The role, or occupation, of a species in a community.

exploitative competition An adverse interaction which results from organisms depleting their shared resources.

family A category, in the classification of evolutionary lineages, of related organisms that ranks above a genus and below an order. A family usually contains many genera.

Grinnellian niche The requirements and behaviors expressed by a species wherever it normally occurs.

Hutchinsonian niche The set of environmental conditions, or opportunities, that will permit a species to exist indefinitely. The set of opportunities that are available to a guild can be referred to as a "nook."

individualism of species Refers to the observation that the distributions and abilities of most species do not exactly coincide because each species is the product of a unique evolutionary history. Coevolution and convergence may occur, but rarely are these processes so complete that groups of species are distributed as a unit or superorganism.

taxon A general term used to indicate groups of related organisms at any level in a taxonomic hierarchy. In the classification of evolutionary lineages, the term can refer to species, genera, families, orders, etc.

INVESTIGATING ENTIRE natural communities is a formidable task because one must take a high diversity of species into account. To focus on a more manageable unit, ecologists usually restrict their attention to some portion of the larger system, such as the "plant" or "bird" community. Such "taxon"-defined communities, however, are rather unwieldy, and they may contain a heterogeneous mix of interactions. Thus, most beetle communities contain predators, herbivores, and scavengers but ignore the bugs, caterpillars, flies, and gastropods that share many of the beetles' resources. The guild concept, which groups species according to the manner in which they exploit a common resource, provides a manageable, functional unit for studying patterns of adaptation and the organization of natural communities. The existence of species with mixed requirements and different evolutionary histories, how-

Encyclopedia of Biodiversity, Volume 3

ever, makes it difficult to determine standard procedures for defining guilds. The indefinite boundaries that surround most guilds are a fundamental outcome of evolution, which is also at play in all of our attempts to group species into communities, trophic levels, and other ecological categories. Discovering how to draw valid conclusions from entities with vague boundaries is a special challenge to ecologists and provides us with a quest of broad significance.

I. THE DEFINITION AND PROPERTIES OF GUILDS

Guilds are groups of species that exploit the same class of resources in a similar way (Root, 1967). This simple definition has several implications that are best explained by way of an example. Consider the birds that probe for insect prey on the bark of tree trunks (Figs. 1 and 2). In many parts of North America this guild is composed of small woodpeckers (family Picidae), nuthatches (family Sittidae), and treecreepers (family Certhiidae). The guild also contains the black-and-white warbler (family Parulidae), whose relatives are primarily adapted for gleaning insects from foliage. Guilds thus contain species with very different phylogenetic histories and, as a consequence, with different propensities for dealing with the special requirements associated with exploiting a particular resource. Thus, the woodpeckers and treecreepers have evolved stiffened tails to act as props in climbing tree trunks, whereas the nuthatches and warbler have evolved modifications of the foot to assist in moving on vertical surfaces.

In addition to the core members of the guild which take the bulk of their food by foraging on trunks and limbs, there are many species, such as titmice (family Paridae), that occasionally feed by probing bark. The presence of these infrequent bark probers illustrates that guilds can have a hierarchical structure with members that range from strict specialists to occasional opportunists. As a consequence, the size and makeup of the guild will depend on where we draw the line with respect to the proportion of a species' diet that is obtained while probing bark.

The hierarchical structure of guilds is also reflected in how broadly we interpret the "in a similar way" portion of the definition. For instance, there are ants that forage for insect prey on tree trunks. These insects have the potential of influencing the food supply available to the bark probers. Their foraging style, however, is fundamentally different: Ants rely on their small size to enter bark crevices, whereas the birds employ their relatively large mass to lever out embedded prey and to flake bark. In this example, it can be seen that the scope of a guild's delimitation needs to be matched to the question being asked. The broader definition, which includes ants, might be most useful if our primary concern is competition, and the narrower concept, which excludes ants, might be more appropriate if our primary interest is in the evolution of adaptations. Furthermore, the birds eat the ants, which introduces further complications if the ants are lumped into the same guild with the birds.

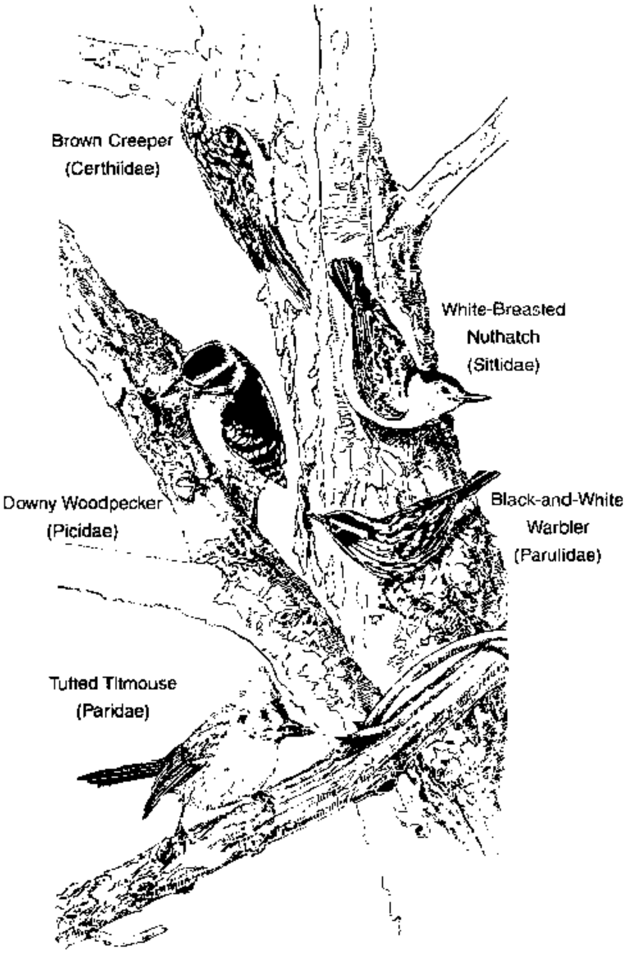


FIGURE 1 A few members of the bark-probing guild in North America.

During the northern winter, the black-and-white warbler migrates to the American tropics, where it joins another representation of the bark-probing guild which includes the tropical woodcreepers (family Dendrocolaptidae) and barbtails (family Furnariidae) shown in

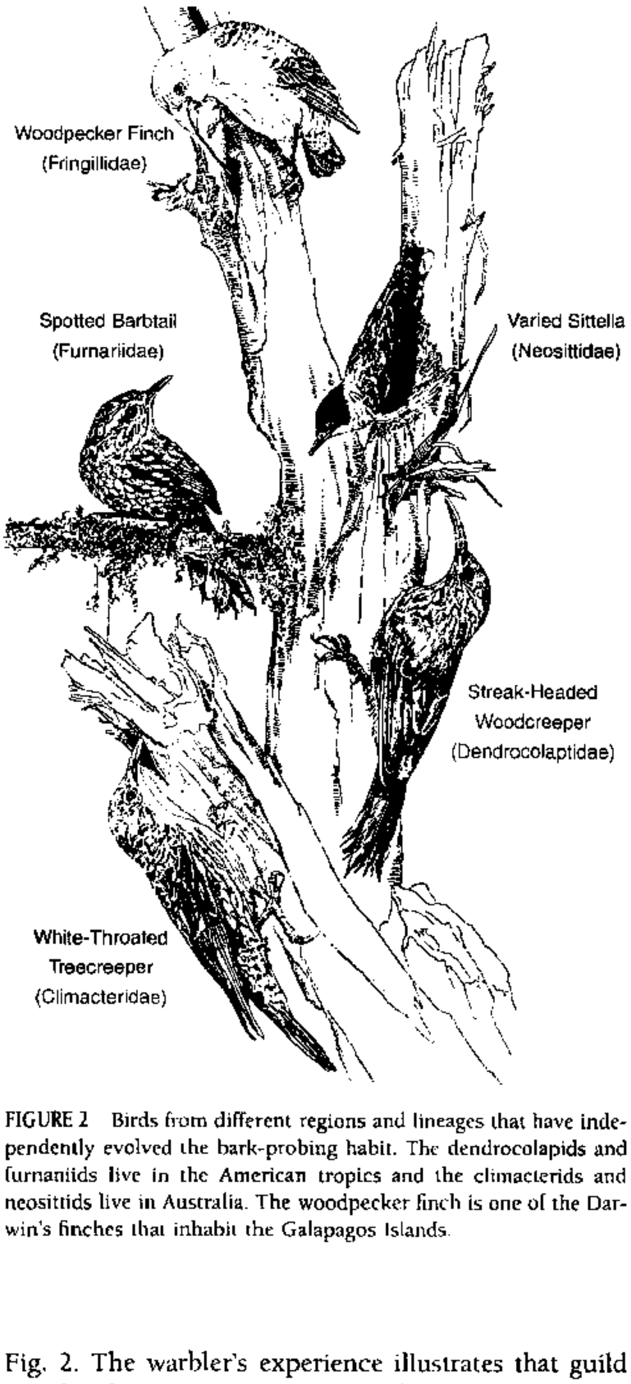


FIGURE 2 Birds from different regions and lineages that have independently evolved the bark-probing habit. The dendrocolapids and furnariids live in the American tropics and the climacterids and neosittids live in Australia. The woodpecker finch is one of the Darwin's finches that inhabit the Galapagos Islands.

Fig. 2. The warbler's experience illustrates that guild membership can vary in space and season.

The opportunities provided by the existence of bark-dwelling arthropods, many of which are in a cryptic resting stage, has been exploited by bark probers that evolved from several phylogenetic lineages throughout the world. Thus, the Australian manifestation of the bark-probing guild contains Australasian treecreepers (family Climacteridae) and sittelas (recently placed in their own family, the Neosittidae). On Madagascar, the bark-probing role is filled by the nuthatch-vanga (family Vangidae), and on the Galapagos Islands the woodpecker finch (family Fringilliade) uses cactus spines to probe for prey under bark.

In classifying the ways that organisms exploit resources, guilds hold a rank that is similar to genera in phylogenetic schemes. One may also think of a guild as a group of species that occupy similar niches. Indeed, one of the original motivations for introducing the guild concept was to clarify confusion regarding the niche concept, which was used to describe three quite different entities: (i) the Hutchinsonian niche, which refers to the set of conditions that are sufficient for a species to exist in a particular habitat; (ii) the Grinnellian niche, which refers to the requirements and behaviors that are expressed wherever a species normally occurs; and (iii) the Eltonian niche, which refers to the role or occupation of a species in a community. The Eltonian niche was usually viewed as a relatively broad category (e.g., sap-feeding insects and predators of small mammals) which could contain several species and occur in a variety of habitats. In creating the guild concept, we provide an alternative category which clarifies some of the ambiguities associated with the Eltonian niche. For instance, by substituting guild for the Eltonian niche, we can avoid the contradictions that occur when several similar species are said to occupy the same "niche," a category that is supposed to be a property of individual species according to the Hutchinsonian and Grinnellian concepts. Accordingly, species exploit niches and guilds exploit "nooks"—the adaptive space that is presented by resources with similar characteristics.

There are many types of guilds in addition to those that are based on food resources. For instance, animals can be grouped according to their use of tree cavities for nest sites and plants can be grouped according to their shared use of agents for pollination (e.g., moths or hummingbirds) and seed dispersal (e.g., fruit-eating birds or wind). In the case of plant-eating insects, there has been a tendency to define "mixed" guilds that group species according to a combination of overlapping functions. Frequently, the "feeding guild," defined on the basis of the insects' mode of feeding and the plant tissues they ingest, is combined with the "sheltering guild," defined on the basis of the insects' use of different plant structures as lairs for protection from enemies and adverse conditions. As a result, we often find that plant-chewing insects are divided into categories such as "leaf-miners," "leaf-rollers," "shoot-borers," and "stem galls" that have been shaped by the interaction of nutritional and protective functions. In searching for patterns, it is important to take all these mixed functions into full account.

It should also be kept in mind that species can belong to more than one guild. For instance, the woodpeckers mentioned previously are members of both the feeding guild that probes bark and the nesting guild that utilizes tree cavities. Many insects shift from one guild to another during the course of development because larvae and adults are adapted to perform such different functions. Thus, most butterfly species have caterpillars that are in the leaf-chewing guild and adults that are in the nectar-drinking guild.

II. THE UTILITY OF GUILDS IN THE SEARCH FOR ORGANIZATION AND PATTERNS

A. Guilds Provide a Natural Unit of Manageable Size for Comparative and Evolutionary Studies of Species Interactions

Even fairly simple communities, such as old fields in the temperate zone, can harbor more than 1500 species of insects belonging to more than 175 families. Such diversity obscures our view of fundamental processes, such as competition, mutualism, and evolution, because most species interact only rarely. This requires that we sort out the insignificant interactions to appreciate the true intensity of the critical interactions. For instance, most species of plant-eating insects in an old field rarely come into direct contact with one another because they are adapted for feeding on different plant species. As a consequence, the use of standard sampling methods, such as sweeping vegetation with a net or collecting the moths that are attracted to lights, which combine the insects on different plant species, would blur our understanding of competitive interactions between plant feeders. Furthermore, the existence of many types of species presents technical difficulties. Thus, to measure the density of all the plant feeders in a community might involve censusing deer, rabbits, mice, sparrows, snails, grasshoppers, caterpillars, aphids, galls, and tiny mites. Clearly, each of these categories of plant feeders require somewhat different methods based on different assumptions, relating to different spatial scales, and having different biases. To avoid these difficulties, ecologists often restrict their attention to some portion of the community. The guild concept is ideally suited to this purpose.

In defining a guild, we create a community detachment whose members share resources that are exploited in a similar way. In many situations, guilds have an advantage over partitioning on the basis of taxonomy because the members of a taxon may be utilizing a variety of resources in a variety of ways. For instance, a bird community may contain rodent-eating owls and seed-eating sparrows but take no account of rodent-eating weasels and seed-eating mice. This is not to imply that guilds are necessarily invalid if all their members are drawn from the same major taxon; frequently, certain styles of exploiting resources are only "available" to organisms with capabilities that evolved in particular taxa. As discussed previously, birds possess the beak and body mass required to probe tree bark. Focusing on a guild has the added advantage of improving our ability to conduct accurate censuses on all the guild members because by sharing the same resources they are likely to be found in the same microhabitats and, by behaving in a similar way, they are more likely to have similar reactions to being observed or captured. Furthermore, the resources, by virtue of belonging to the same class, can be more accurately measured and compared. As a consequence, we are in a position to draw closer comparisons and conduct more realistic experiments when we are working with a guild.

By observing how a guild changes along spatial gradients, we can seek patterns in the relationships between environmental factors and variations in population density, reproductive rate, body size, dispersal ability, investment in predator or herbivore defenses, and a variety of other traits within a group of species that are comparable because they use similar resources in a similar way. Such comparisons also permit us to examine how species are "packed" onto the guild's resource base in relation to disturbance, elevation, and latitude.

Comparisons of similar guilds can reveal interesting patterns. For instance, there is a much smaller range of sizes of plant-chewing insects that live in abodes, such as mines and leaf rolls, compared to that of plant chewers that feed in exposed locations. This pattern reflects the limitations on body size that are imposed by the small spaces provided by plant structures that are pliant enough to be penetrated or rolled.

Much can be learned by comparing the same guild in regions that have been isolated from one another so that their biotas have undergone independent evolution (Figs. 1 and 2). In such comparisons, the discovery of close resemblance provides insights into the process of convergent evolution. In particular, we can observe (i) if certain types of characters are more responsive to evolutionary pressures, (ii) if certain combinations of traits are associated with each other in different regions, and (iii) the degree to which convergence is limited by deep ancestral traits in the various phylogenetic lines that make up the guild.

B. Guilds Partition Biotas into Entities Appropriate for Detecting Community Organization

Ecologists have long debated the significance of exploitative competition in organizing communities. An important line of inquiry in this debate concerns the constancy of a community's functional relationships. The reasoning is as follows: If communities are competitively organized, we expect to find that important functional relationships are predictable in space and time because the waxing and waning of populations of one species would be countered by compensatory changes in the populations of other species that use the same resource. Furthermore, these compensatory responses are expected to be most intense within groups of species that have similar methods of using the shared resources. Finally, we might expect to find that similar types of communities have similar functional profiles because the imperatives associated with using similar resources have sorted and shaped species into convergent groups. Guilds provide a means of partitioning a community into functional entities in which these compensations might be detected.

Much of the search for community organization has focused on plant–insect associations. One reason for doing this is that the basic resources (the biomass of foliage, stems, roots, flowers, and seeds produced by a particular plant taxon) are well circumscribed and relatively easy to measure. The constancy of the association's functional profile can be judged by changes in the guild spectrum—the proportion of total herbivore biomass that is engaged in exploiting different plant parts in various ways. A matrix for sorting insect herbivores into guilds is presented in Table I. In all likelihood, such a scheme will need to be revised to suit the needs of a particular investigation.

Evidence for community organization based on guilds has been elusive. For instance, Cornell and Kahn found extensive variation in the guild spectra associated with 28 tree species in Great Britain and Root and Cappuccino found no evidence for compensatory changes in the densities of species within guilds that feed on goldenrods. These failures draw us to question our assumption that the intensity of competition is likely to be most severe within guilds. Different types of resources are linked in nature. Thus, in plant–insect associations, the consumption of leaves will reduce the future availability of seeds and the sapping of juices from stems will reduce the nutritional quality of foliage eaten by leaf chewers. Similarly, the caterpillar eaten by a foliage-gleaning bird is unavailable as a moth to bats that forage in the air. Linked resources link guilds and open the possibility for exploitative competition to play out its effects across a diffuse network of species. In other words, the use of a common resource seems to be of primary importance in competitive interactions; the manner in which the resource is exploited is of little consequence unless differences in a guild's behavior permit members to use an exclusive subset of resources. As a consequence, there are many cases of competition between guilds. For example, Schluter

TABLE I

A Provisional Martix for Defining Guilds of Insects that Feed on Terrestrial Plants

Manner of feeding	Resource Buds and leaves	Pliant stems	Flowers	Pollen	Fruits	Seeds	Wood and bark	Crusts of algae, molds, etc.
Chewers								
Exposed	✓	✓	✓	✓	✓	✓		✓
Concealed	✓	✓	✓		✓		✓	
Sap-tappers	✓	✓	✓		✓	✓		
Gall-makers	✓	✓	✓		✓			
Grazers				✓				✓

found that Darwin's finches in the Galapagos Archipelago consumed more nectar on islands on which nectar-feeding carpenter bees were absent.

Of course, competition could be operating in a diffuse fashion that would be difficult to detect by only studying variations in guild spectra. In plant–insect associations, however, there is increasing evidence from long-term experiments and extensive surveys that plants in natural communities are rarely devastated by herbivores and that competition between plant-eating insects is a sporadic occurrence. These results, considered in conjunction with the highly variable guild spectra that have been reported by Cornell, Root, and Strong, suggest that most plants could support more herbivore species (i.e., plants provide Hutchinsonian niches that have not been filled) and that the structure of these associations is idiosyncratic, largely determined by the characteristics of a few dominant species that happen, for a variety of historical reasons, to have evolved to become members of the community. Many ecologists have the impression that guilds and communities that are dominated by predators or vertebrates are more likely to be organized by competition and to exhibit more predictable functional profiles. This impression, however, has not been adequately evaluated.

C. Guilds Provide a Framework for Describing and Comparing the Trophic Structure of Ecosystems

The trophic structure of ecosystems is often described in terms of broadly defined levels consisting of primary producers, primary and secondary consumers, decomposers, and so on. Food webs, on the other hand, are described in terms of individual species. Various attempts have been made to develop a scheme based on guilds that can be used to describe trophic structure with an intermediate degree of refinement; thus, herbivores might be subdivided into browsers of woody plants, grazers of grasses, sap-tappers on succulent foliage, borers in stems, and so on. To accomplish this requires that the entire community be partitioned so that all its members can be placed into a standard set of feeding guilds that have a widespread occurrence in nature. When ordered in this way, it is hoped that insights about energy flows can be gained by drawing comparisons between systems. In practice, however, such schemes have been little used, probably because it is almost impossible to develop a "key" to the guilds of the world that is flexible enough to accommodate a wide array of taxa and yet sufficiently circumscribed that different ecologists will assign species to guilds in the same manner.

III. THE INDIVIDUALISTIC NATURE OF SPECIES AND THE VAGUE BOUNDARIES OF ECOLOGICAL CATEGORIES

The process of defining guilds raises some fundamental questions. Are guilds distinct entities that reflect the outcome of natural processes or are they merely artificial groupings that we invent to divide diversity into more comprehensible units? In other words, are the opportunities provided by the environment discontinuous so that species are sorted into clearly defined groups on the basis of their requirements and lifestyles? Furthermore, are similar openings (or nooks) available in different regions? Are organisms constrained from filling these opportunities by their evolutionary histories? We can begin to address these difficult questions by comparing the characteristics of the core and marginal members of various guilds.

If we census the habits of all the species that utilize a particular resource, we usually find that there is a wide variation in the level of their dependency on the resource. The most dependent species—the core users—usually possess obvious specializations. Thus, in the bark-probing guild (Fig. 1), the treecreepers, woodpeckers, and nuthatches, which take the bulk of their diet from bark, have specialized feet and tails for moving on vertical surfaces (Richardson, 1942). In our census, however, we usually encounter a host of species that use the resource only rarely. These marginal users are either generalists or they are specialized along other lines that compromise their ability to use the resource in question. For instance, bark-probing titmice often forage on horizontal surfaces, such as the tops of limbs and branches, where they take much of their prey by hammering apart acorns, galls, and similar objects that they hold against the perch with their foot. These habits are reflected in the distinctive foraging maneuvers that titmice display during their infrequent forays onto tree trunks. Other species, such as gnatcatchers, vireos, and wood warblers, which are primarily adapted for gleaning insects from tree foliage, make clumsy efforts to take insects from bark on rare occasions.

Here we confront the individualistic nature of species—a fundamental issue, originally raised by Gleason, that complicates all of our efforts to define functional groups of species. Thus, the various taxa that can utilize a particular resource have evolved "individualistically" along independent paths. As a result of their separate histories, these taxa have different constraints and proclivities, which result in different levels and styles of specialization. As a consequence, guilds often have ambiguous boundaries consisting of several species of generalists and inept opportunists that are specialists on other resources. Since the degrees of dependency grade into one another, the line one draws to define the membership in a guild can be somewhat arbitrary.

The indefinite boundaries of guilds are merely one expression of a more general issue confronting ecologists. Community ecology and biogeography are, by their very nature, concerned with levels of organization that consist of multiple species. In their efforts to discover the processes operating at these levels, workers have developed several systems for classifying species into groups on the basis of similarities in their (i) response to physical or site conditions to define communities; (ii) possession of particular adaptive traits to define life-forms and adaptive syndromes; (iii) response to seasonal cues to define phenological aspects; (iv) geographic distributions to define biogeographic provinces, biomes, life zones, and plant formations; and (v) diet to define trophic levels. Ecologists recognize and discuss these entities because they find that several species fall into clusters on the basis of these various types of classification. As with guilds, however, we also encounter species with intermediate characteristics that tend to blur the boundaries of these various categories. This is because all the traits that underlie these classifications are subject to the same individualistic evolution that can produce the marginal members of a guild. Thus, vague boundaries are an inherent property of any species assemblage—a fact that requires ecologists to cultivate certain habits of mind in forming arguments.

IV. ATTITUDES CONCERNING VAGUELY BOUNDED CATEGORIES

People are generally uneasy about categories, such as guilds, that are subjectively defined because the criteria one chooses to define the operational boundaries will have an impact on our ability to observe patterns. Thus, if a guild is defined too broadly, the characteristics of a miscellaneous collection of marginal species could obscure interesting similarities between the core members. On the other hand, if a guild is defined too narrowly, we could overlook the full range of influences that stem from extracting a particular resource in similar ways. There have been a variety of attempts to address this problem by using quantitative procedures, such as cluster analysis and principal components analysis, to delimit the membership in guilds. The raw material for these analyses comes from "activity censuses" which measure the relative frequency that species utilize different resources in different ways. The groupings that are sorted out by these procedures, however, often seem abstract and artificial. Furthermore, they are probably no less arbitrary than those that are more subjectively defined because the person who is running the analysis must decide how broadly to cluster the species' activities.

Despite these difficulties, the continued and widespread use of subjectively defined categories attests to the need of such concepts in ecology. The best results are obtained when we follow some simple guidelines. In the case of guilds, the classifications that provide the most insights are based on a thorough knowledge of the species' natural history. Open-minded observations and journal keeping are good starting points to form an intuition for distinguishing resources with properties that require special means for their efficient exploitation. For instance, insects on bark constitute an appropriate resource for a guild of bird-sized predators because the core members must move on vertical surfaces and extract prey from narrow crevices or under bark—maneuvers that are difficult for birds that lack the necessary specializations.

After a guild has been tentatively defined, it helps to conduct activity censuses at different sites and in different seasons to discipline intuition and provide a quantitative basis for determining guild membership. (For sedentary organisms such as plants, one would count the incidence of the traits that define the guild along transects laid out in appropriate habitats.) When the investigators are satisfied that a guild "makes sense" with respect to the questions that are being asked, the criteria that will be used to delimit the boundaries should be clearly described and fully justified. At this point, it may be useful to assign classes of membership; for instance, core species might be those that engage in the activities that define the guild on at least 50% of occasions, and marginal or accidental species might be those that engage in such activities on less than 10% of occasions. Many of these steps seem obvious, but they are often left out of guild classifications, especially in cases in which the investigators are attempting to partition entire communities in such a way that each species can be assigned to a single guild. Such classifications are often used to compare the functional organization of communities in which only the most general of categories can accommodate the great diversity of species that must be placed.

Part of the art of becoming an ecologist involves developing a set of attitudes for coping with the complications that stem from the individualistic nature of species. One of the most important of these is an ability to match the question one is asking with the most appropriate grouping of species—the set that will reveal valid patterns that act in nature. Thus, the guild definition that is most useful for exploring convergent evolution may be quite different from the one that is best for observing compensatory shifts in the densities of interspecific competitors. In addition, we need to be on steady guard against the natural tendency to drift into thinking that the entities we have invented are "real." Guilds are not a fixed feature of nature. They are a convenience that we can use to cope with diversity, reveal patterns, and facilitate discussion. As a consequence, guilds can be modified as long as they provide a valid base for addressing a question and their limits are explicitly defined and justified.

See Also the Following Articles

COMPETITION, INTERSPECIFIC • FUNCTIONAL GROUPS • HABITAT AND NICHE, CONCEPT OF • TROPHIC LEVELS

Bibliography

Cornell, H. V., and Kahn, D. M. (1989). Guild structure in the British arboreal arthropods: Is it stable and predictable? *J. Anim. Ecol.* 58, 1003–1020.

Curtis, J. T. (1959). *The Vegetation of Wisconsin.* Univ. of Wisconsin Press, Madison.

Gleason, H. (1926). The individualistic concept of the plant association. *Bull. Torrey Botanical Club* 53, 7–26.

Hawkins, C. P., and MacMahon, J. A. (1989). Guilds: The multiple meanings of a concept. *Annu. Rev. Entomol.* 34, 423–451.

Richardson, F. (1942). Adaptive modifications for tree-trunk foraging in birds. *Univ. California Publ. Zool.* 46, 317–368.

Root, R. B. (1967). The niche exploitation pattern of the blue-gray gnatcatcher. *Ecol. Monogr.* 37, 317–350.

Root, R. B., and Cappuccino, N. (1992). Patterns in population change and the organization of the insect community associated with goldenrods. *Ecol. Monogr.* 62, 393–420.

Simberloff, D., and Dayan, T. (1991). The guild concept and the structure of ecological communities. *Annu. Rev. Ecol. Syst.* 22, 115–143.

Strong, D. R., Lawton, J. H., and Southwood, R. (1984). *Insects on Plants: Community Patterns and Mechanisms.* Harvard Univ. Press, Cambridge, MA.

Wiens, J. A. (1989). *The Ecology of Bird Communities.* Cambridge Univ. Press, Cambridge, UK.

HABITAT AND NICHE, CONCEPT OF

Kenneth Petren
University of Cincinnati

GLOSSARY

community An ecological term referring to a set of species that occur in the same location that have the potential to affect each other either directly or indirectly.

community structure The web of potential biological interactions among members of a community that may be characterized in terms of diversity, complexity, hierarchy, and stability.

ecotone A zone of transition between two different habitats that may contain a community of organisms distinct from either habitat.

habitat structure Analogous to community structure, but limited to the physical structural aspects of a habitat. The structure of habitats may be characterized by such measures as complexity, heterogeneity, regularity, stratification, and fractal dimensionality.

microhabitat Locations within a habitat where organisms may carry out important aspects of their lives, such as places for harvesting food, nesting, or taking shelter.

niche overlap The proportion of available resources that are shared by two species. Usually used in the context of a single resource that limits population growth.

WHY ARE SPECIES COMMON in some places but rare or absent in other places? Early attempts to answer this fundamental question led to the concept of the habitat. A habitat is simply where an organism can be found in nature. Habitats are described in terms of geography, geology, climate, as well as by other species commonly found within the same habitat. The habitat concept is very closely related to another concept used to characterize species, the niche. The niche of an organism can best be described as its role in the community of organisms around it. Among the many traits that may comprise a niche are the physical and climatic characteristics that an organism can tolerate, and the resources required for the species to persist, such as food and shelter. The habitat of a species is often treated as a subset of its niche, and in practice it has grown difficult to draw a distinct line between the two concepts, yet the study of habitats has a history distinct from that of the niche. Today, habitat descriptions are commonly used as a practical guide for locating and maintaining species, whereas the niche is a more abstract concept that forms the conceptual foundation for much ecological and evolutionary theory.

Encyclopedia of Biodiversity, Volume 3

I. CONTRASTING THE HABITAT WITH THE NICHE

As is the case with many ecological concepts, defining habitats and niches concisely and unambiguously is difficult. Habitat definitions are particularly prone to problems of scale. At the larger end of the scale is the biogeographical term *biome*. Some biomes may be considered the habitat of larger, well-traveled species such as large birds of prey, but usually the habitat is defined more narrowly. On the smaller end of the scale, the term microhabitat is used to describe the places where an organism spends part of its time. For instance, a fish may forage in the microhabitat that occurs near the banks of a river. Thus the concept of the habitat lies somewhere in between the biome and the microhabitat.

An ecological community is a suite of species that occur in the same location that at least have the potential to interact and affect each another either directly or indirectly. These effects are usually measured in terms of population growth or changes in the density of individuals. Different communities of organisms will generally reside in different habitats. If most species in one community have little chance of affecting the populations of other species in a different community, then these communities probably occur in different habitats. In this way, microhabitats refer to places within a community, and biomes encompass a number of different habitat types. Definition of habitat and community boundaries will depend on the exact species under consideration. For instance, it may be reasonable to distinguish the small mammal communities of grassland from nearby riparian forest, though an individual jaguar may prey on species in both habitats.

The habitat concept was originally applied to single species, but because many species share similar habitats, some general habitat descriptions are applied to many species. Habitat descriptions used in this fashion often incorporate the dominant species commonly found in that habitat. In nature, borders between different kinds of habitats are often not very distinct. Generally, one habitat type will gradually give way to another creating a zone of transition. Communities in the transition zone may be qualitatively different from communities that occur within one habitat type or another.

The primary habitat preferences of a species can be displayed graphically. Ordination is the process of characterizing a species with regard to habitat gradients. Gradients can be displayed as axes on a graph. Usually, some measure of density or population growth is plotted, and the portion of the gradient that is occupied by the species is referred to as the habitat breadth of the species. Figure 1 shows how two species of plankton can be compared according to a single habitat dimension, depth in the water column. There are potentially limitless gradients or factors that may be involved in characterizing the habitat of a species. Graphical representation of more than three variables in a single graph is difficult, but the concept easily extends into multiple dimensions.

The niche is a more abstract term than the habitat. It encompasses all possible interactions that a species has with the environment and other species in the community. Conceptually, the niche is richer than the habitat and forms the foundation for much ecological theory. Because of the abstract nature of the concept, the definition is perhaps even more difficult to formulate for all species. There are striking similarities in the way niches and habitats are displayed. Like a habitat, a niche is often defined by axes, and there are multiple axes that can be compared.

The most common type of axes used to characterize a niche include those that define environmental tolerances and resource requirements. Humidity, sunlight, temperature, wind exposure, and pH are examples of environmental axes that help to define a niche. Resource axes may include food resources (insects, seeds, bacteria, nutrients), space requirements (breeding sites, ref-

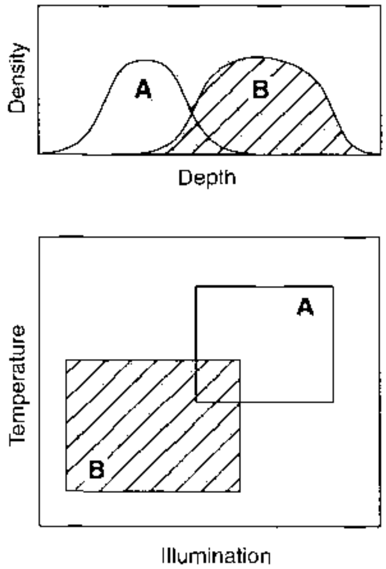


FIGURE 1 Two hypothetical species of plankton compared along the habitat axis of depth in the water column (top panel). This single habitat axis is correlated with two different niche axes, temperature and illumination. Species B is found in shallower, brighter, and warmer areas.

uges, foraging zones). Figure 1 also compares the two plankton species in two niche dimensions, illumination and temperature, which are correlated with the habitat axis, depth. This highlights the similarity of habitat and niche axes, which are often difficult to distinguish clearly.

In practice, ecologists focus on a subset of niche dimensions that are the most important in determining the role of a species in the community. Among these are food resources that are essential for survival. These factors *limit* the population size of the species and therefore determine the distribution and abundance of the population.

In spite of the more limited role of the habitat in theory, the physical nature of habitat axes make them useful tools for practical purposes. Habitats can often be adequately described with relatively few variables, whereas to fully document the niche of a species requires careful study and a great deal of time. Yet if the habitat can be accurately captured, it is likely that many the complex niche requirements and interconnections will be contained within the same habitat. As stated earlier, communities tend to map one to one onto habitats. The utility of this relationship is exploited for management when there is insufficient time to study all aspects of an organism's ecology before an action needs to be taken. To preserve a single species, one must also preserve its niche requirements and the complex interconnections with the rest of the community. Preserving adequate habitat will often achieve this goal. This principle is important not only for managing target species, but for preserving and restoring biodiversity in general.

The primary habitat and niche for a number of species is given in Table I. These brief characterizations are typically used as a shorthand way to describe where to find species and the basic role played by each in the community. Habitat descriptions include geographical information and often refer to the dominant species present (e.g., grassland, coral reef). Niche descriptions for a species usually begin with the main food items consumed.

As one moves through this list of very diverse species, the differences in the scales and roles are readily apparent. Trees may define different successional stages of a forest, each of which may be considered a different habitat. A jaguar may routinely cross into different successional stages on a daily basis, making it difficult to clearly define a jaguar's habitat. In contrast, certain trees, birds, and lichen are limited to only a certain successional stage, while specialized insects, parasites, and pathogens may even be limited to a single species of host. These differences make broad comparisons of niches and habitats difficult, but at any single scale, these concepts can be applied to facilitate meaningful comparisons among similar species.

II. HISTORY OF THE HABITAT AND THE NICHE

A. Niche Origins

It is tempting to think of a niche as a physical place. This is the common usage of the word, and there are examples in the early ecological literature that use the term *niche* in the purely physical sense. However the origins of the ecological niche reside in the more general observation that no two species are exactly alike.

Many naturalists in the latter part of the 19th century turned their attention toward documenting the traits that distinguish one species from another. When two species appeared very similar, it was thought that eventually differences could be found that would distinguish the unique role of each in the community. This idea was evident even in the writings of Darwin, and over time it has evolved into a very general principle: the principle of competitive exclusion. Exploring the implications of this relatively simple notion has dominated the study of ecology for much of the 20th century.

The principle of competitive exclusion and the concept of the niche developed in parallel through the early 1900s. Theoretical work by Volterra (1926) showed that if certain assumptions are met, only one species should be able to survive on a single resource. Gause (1934) demonstrated this principle experimentally with two species of *Paramecium* feeding on a common resource. Two species that consume the same resource in the same way simply cannot coexist.

The term niche was first used in the context of competitive exclusion by Grinnell, a superb naturalist of the North American west, as early as 1914. Grinnell's study of the California thrasher represents one of the first that specifically set out to characterize the ecological niche of a species, though he laid out the concepts of the niche and competitive exclusion 10 years earlier in a study on chickadees.

Grinnell's study of the California thrasher illustrates some of the most basic components of a niche (a) type of food consumed (mostly insects, berries at some times of the year), (b) microhabitat preference (beneath shrubby vegetation), (c) physical traits and behaviors used in gathering food (a long beak thrashed through the top layers of soil and leaves), and (d) resources

TABLE I

The Primary Habitat and Major Niche Components of Some Representative Species

Species	Habitat/Biome	Niche
California thrasher *Toxostoma redivivium*	Coastal chaparral; western North America	Generalist insectivore; territorial; sometimes a fruigivore
Cuban crown giant lizard *Anolis equestris*	Tropical forest tree crowns; Caribbean (Cuba)	Generalist insectivore; short pursuit predator; territorial
Spotted salamander *Ambystoma maculatum*	Ephemeral ponds in deciduous forest; temperate North America	Generalist carnivore, breeds in ponds with few large fish
Jaguar *Panthera onca*	Tropical forests of all types, savanna; Central and South America	Small animal carnivore; solitary stalking predator, keystone predator
Plains zebra *Equus burchelli*	Tropical savanna, Africa	Grazer of short grass; gregarious; migratory
Harvester ant *Pogonomyrmex rugosus*	Sandy regions of deserts; southwest North America	Scavenger of small seeds; makes terrestrial, colonial burrows
Tawny mining bee *Andrena fulva*	Temperate fields, gardens; Europe	Nectar/pollen harvester; solitary soil burrower; pollinator
Camel cricket *Ceuthophilus silvestris*	Caves, moist environments; North America	Scavenger, detrivore
Ghost crab *Ocypode quadrata*	Sandy marine beaches; temperate and tropical Nearctic	Scavenger; territorial; nocturnal
Mussel *Mytilus californianus*	Intertidal, rocky marine shores; western temperate North America	Filter feeds on Plankton; sessile; pelagic larvae
Tardigrade (water bear) *Echiniscus spp.*	Moist film on plants, moss, or lichen; cosmopolitan	Plant and moss cell predator; meiofaunal (0.4–1.0 mm)
Guanacaste tree *Enterolobium cyclocarpum*	Tropical savanna, dry forest; Central America	Large canopy tree; pioneer, keystone species
Flowering dogwood *Cornus florida*	Temperate deciduous forest, North America	Shade-tolerant, understory; pollen and nectar producer
Deuteromycota fungi *Arthrobotrys spp.*	Moist soil, pools; cosmopolitan	Sit-and-wait predator of roundworms, hoop-snare structures
Slime mold *Acrasiae spp.*	Moist soils, various habitats; cosmopolitan	Bacteria consumer, single and multicell stages
Schistosome	Humans and freshwater	Blood parasite in humans, gut
Schistosoma mansoni	Snails; tropical, Africa, South America	Parasite in snails

required for shelter and breeding (dense shrubs for night roosting and nesting). These four basic factors allow one to characterize the basic niche of most animals, and most animals differ with respect to one or more of these factors.

An analogous set of core niche dimensions may be constructed for plants, which tend to partition niche space along resource axes such as available light, soil moisture, and various soil nutrient gradients. Important niche dimensions for marine organisms may include temperature, substrate type, salinity, pH, and exposure to waves.

Thirteen years after Grinnell, the publication of Elton's *Animal Ecology* text (1927) established the term niche in the lexicon of ecology. However it was not until the work of Hutchinson (1957) that the concept of the niche was fully developed as a cornerstone of ecological theory. Hutchinson's niche is an *n*-dimensional hypervolume, implying that there are usually many factors, or dimensions, that define the role of a species in a given habitat. These included yet extended well beyond the basic niche dimensions described by Grinnell.

Other ecological concepts emerged from this foundation. The resources that a species would use if it were isolated from all potential competitors is part of its fundamental niche. The realized niche is a subset of the fundamental niche that includes the resources actually consumed by the species in nature. The difference between the realized and fundamental niche can be attributed to competition from other species in the community.

B. Habitat Origins

Originally, the habitat of a species was not viewed as a part of its niche. In the early history of these terms they were treated separately. The habitat was seen more as a guide to the kind of community in which the organism played its role. For instance, the California thrasher occupies coastal chaparral of western North America, but this habitat description can be viewed as distinct from its niche within that habitat. Also, species were viewed as possessing suites of traits and physical tolerances that enabled them to be suitable only for specific habitats.

Over time, habitat axes were compared right alongside common niche axes, and the distinction became obscured. This change may be due in part to parallel development in the field of evolution and adaptation and increased knowledge of the processes involved. Today it is not difficult to imagine that two species may have evolved to occupy different habitats, whereas at a previous time only one species may have occupied both habitats.

The history of the habitat concept can also be traced back to the work of Grinnell and his contemporaries. In one of the earlier studies, Grinnell and Storer characterized the animals of Yosemite according to the elevations at which they occurred. In this mountainous region, species tend to be found in narrow elevational zones or bands.

At about the same time as the Yosemite study, Ramensky wrote that each species was unique and possessed adaptations that enabled it to tolerate a unique set of environmental conditions. This "principle of species individuality" bears strong resemblance to the early influences on the niche concept, except that in the habitat context, competitive exclusion was not at the core. Instead, species were seen to be adapted to certain conditions and habitats, and they independently sorted themselves out accordingly.

Subsequent work has confirmed that habitat boundaries are often determined by physical tolerances. However there are numerous examples where competition and other species interactions play an important role in determining the habitat breadth of a species in nature. This further illustrates how the habitat and the niche have converged over time. Therefore it is appropriate to treat the habitat and the associated gradients and axes as a subset of the niche.

III. COMPARING HABITATS AND NICHES

A. Multispecies Comparisons

The habitat and niche of any species can be characterized independently, but the true value of these concepts lies in comparing multiple species. The ability to display habitats and niches graphically greatly facilitates comparisons. Figure 2 shows how species can be ordered along a habitat gradient. Copepod species show different patterns of abundance with respect to the gradient from ocean to land in the intertidal region. While in this instance there is a large amount of overlap, each species has a characteristic distribution in this region.

Habitat boundaries are rarely sharply defined. *Ecotone* is the term used to refer to this zone of transition between habitats. Because this zone is somewhat different from either neighboring habitat, an ecotone may contain several unique species. This concept was proposed to help explain distributions of species that did not appear to coincide with habitat boundaries. In a unique test of this idea, Terborgh and Weske (1975) found little evidence that subsets of bird species along an elevational gradient were responding to ecotones. Instead, it appeared that the elevational ranges of many species were independently defined by the presence of other closely related species. In this case, competition appeared to be more important than habitat gradients for setting the actual limits of species distributions.

The niche framework is typically used to compare groups of closely related species that consumed similar resources but differed in one or a few important ways. Suites of species in a community that utilize the same general resource were termed guilds by Root (1967). Just as a guild of blacksmiths all work with iron yet have different specialties, assemblages of animals that specialize on seeds, fruits, or insects all have different ways of harvesting these resources.

Early studies focused on guilds that differed mainly with respect to body size. Hutchinson (1959) observed that species within a guild often differed by a minimum size ratio. Known as Hutchinsonian ratios, these regular arrangements of species body sizes occur in a number

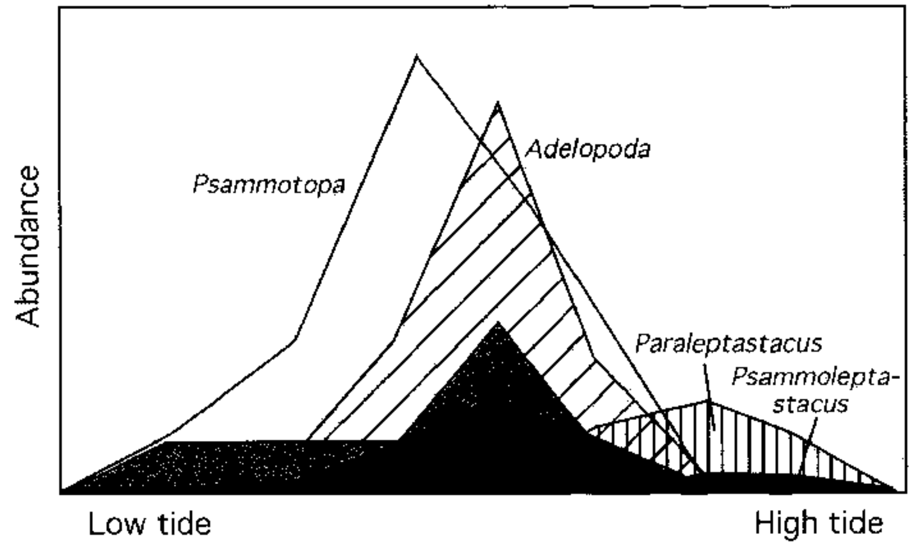


FIGURE 2 The distribution of copepod species along the sandy intertidal habitat gradient in Massachusetts (only genus names are given). From Pennak, 1951.

of vastly different communities consuming very different resources.

One example of the kind of pattern Hutchinson was attempting to explain is the assemblage of rodents that consume seeds in the deserts of North America. Figure 3 shows how similar assemblages of rodents in different communities contain a similar set of rodents. Each species differs from others in the same community mainly in terms of body size. Convergence is a term applied to organisms or communities that are not closely related historically that have members with similar physical features because they occupy similar niches. The implication is that parallel evolutionary trajectories were followed independently by different species because of similar selection pressures.

The underlying mechanisms that create such regular patterns can be explained because each species consumes resources of a size class that is proportional to its body size. Smaller rodents consume smaller seeds, while larger rodents eat larger seeds. In a different community, smaller fruit pigeons consume smaller fruits, while larger species primarily consume larger fruit. For a medium sized bird, small fruits take too long to gather and handle, and the bird can not eat them efficiently enough to sustain itself. Larger fruits may be too big to handle or crack open.

Many types of food resources such as seeds, fruit, or arthropods have a wide, flat distribution along a size gradient. These relationships can be displayed graphically with a single niche axis, food resource size (Figure 4). One curve represents the available resources, in this case seeds, and the narrower curves underneath represent the sizes and amounts of seeds consumed by each species. This division of available resources among members of a community is termed resource partitioning.

Utilization curves are often bell shaped (normal, Gaussian) to reflect the tapering ability of each species to consume seeds away from the "optimum" size. This type of niche representation allows direct comparison of the amount of overlap in resource consumption. If food is indeed the primary limiting resource, the overlap of species resource utilization curves directly represents the amount of competition that each species is experiencing. It is the regular spacing of curves to minimize overlap, or competition, that creates the pattern of regular size distributions of animals seen in natural communities. This simple framework forms the basis of niche theory.

The concepts of fundamental and realized niches can be clearly visualized within this framework of graphical representation of both habitat and niche dimensions. Mussels are capable of living in a wide range of elevations above the low tide mark, and they will move into these spaces in the absence of other species. However other animals, such as barnacles, are better adapted to occupy lower positions and can outcompete mussels in some zones. That is why in natural communities, the realized niche of mussels in terms of this habitat gradient is usually narrower than the fundamental niche.

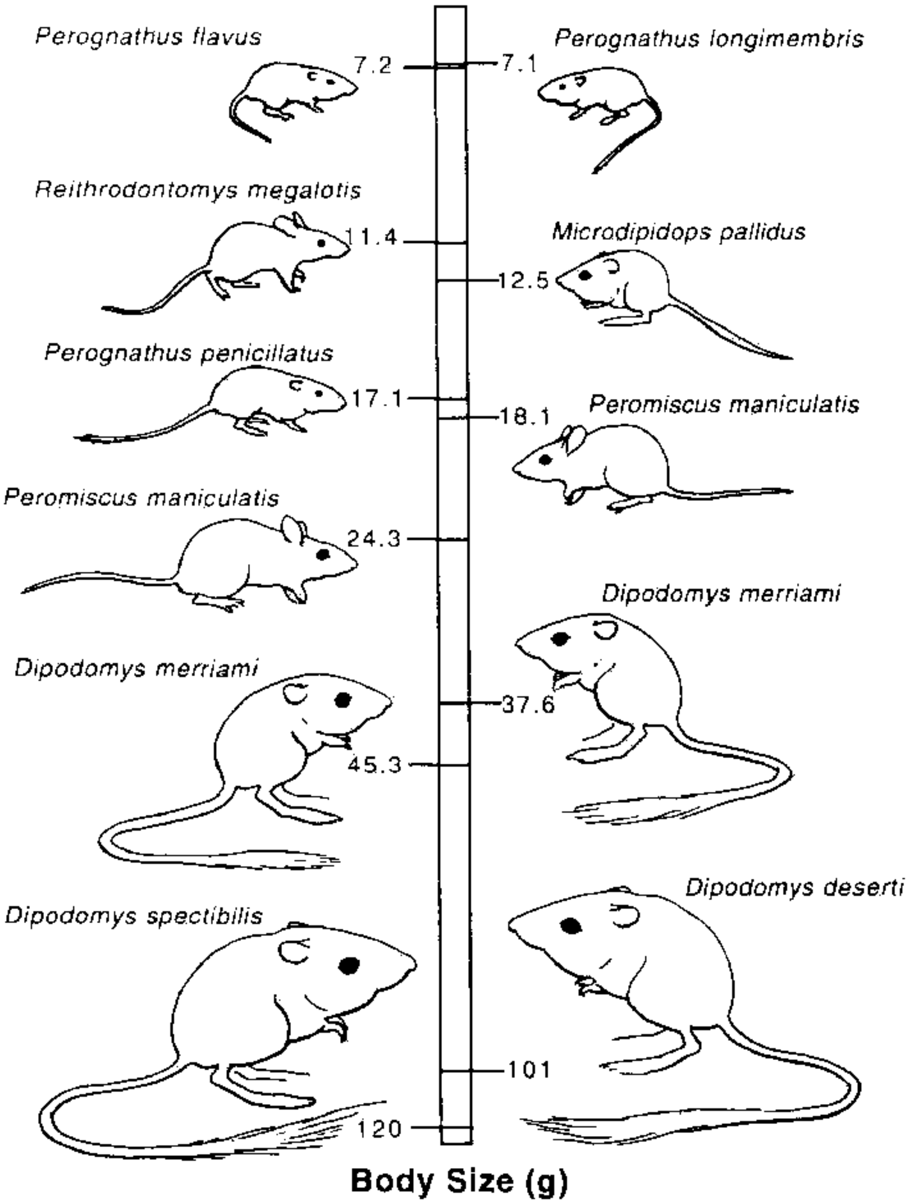


FIGURE 3 A comparison of the niche relations among Great Basin and Sonoran desert rodent guilds. From Brown, 1975. Different habitats support a different assemblage of rodents that show similar differences in body size.

A similar process may be operating in the granivorous rodent example presented earlier. Each rodent may be capable of consuming a wider range of seeds, albeit inefficiently. In nature, either there are fewer of these suboptimal resources available because of competitors, or each species deliberately chooses a narrower range of seeds to consume to avoid competition.

The effect of competition on realized niches can be seen by comparing island and mainland communities. Islands usually support fewer species than equivalent mainland habitats. It has been shown for many kinds of species that in these species-poor island communities, species tend to eat a wider array of foods and therefore have wider niches. This agrees with the expectation based on niche theory outlined earlier: there are fewer potential competitors on islands, and therefore there is less resource overlap with other species acting to constrain their niches.

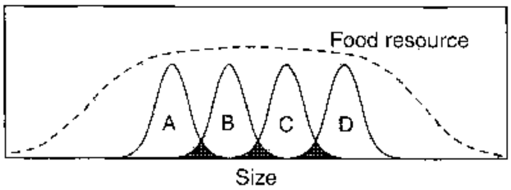


FIGURE 4 A hypothetical species assemblage. A nearly continuous spectrum of resources that differ mainly in size (top curve) are consumed and partitioned among different species that also differ mainly in size (A–D).

Species that have very high overlap in many niche dimensions often have little or no overlap in one key dimension. This principle, called niche complementarity, lies at the heart of the guild concept. In previous examples, similar species consumed the same general resource and had high niche overlap, but they showed little overlap with respect to a single dimension: body size. Typical niche axes where complementary species can be found include size, light (nocturnal/diurnal), benthic/pelagic, and canopy/ground or coniferous/deciduous microhabitat.

In many instances niche differences among species are clear, but sometimes niche differences are extremely subtle and require careful study to be revealed. For instance, five species of *Dendroica* warblers can all be found in the very same tree in temperate zone coniferous forest habitat. All glean similar types of prey, primarily lepidopteran larvae, from the surfaces of the tree. Only careful study by MacArthur (1958) revealed that each species favors different parts of the tree, and each tends to move about the tree in different ways (Figure 5). Thus it is not only the type of food resource that is limiting the population growth of each species, but where and how that food resource was harvested. There are enough subtle differences in microhabitat use and behavior to allow coexistence of species that at first glance appear to share equally the same limiting resource.

Niches can change through time. For instance, the larval forms of many organisms often have completely different niches from adult forms. Thus, the term ontogenetic niche has been applied to amphibians, arthropods, and a wide variety of marine organisms. The great versatility of the niche concept allows incorporation of these temporal changes by simply adding a temporal axis to the species niche. The realized niches of species in a community can also change through time if the community undergoes change. Extinctions and invasions cause rapid community changes that in turn result in niche shifts among other community members.

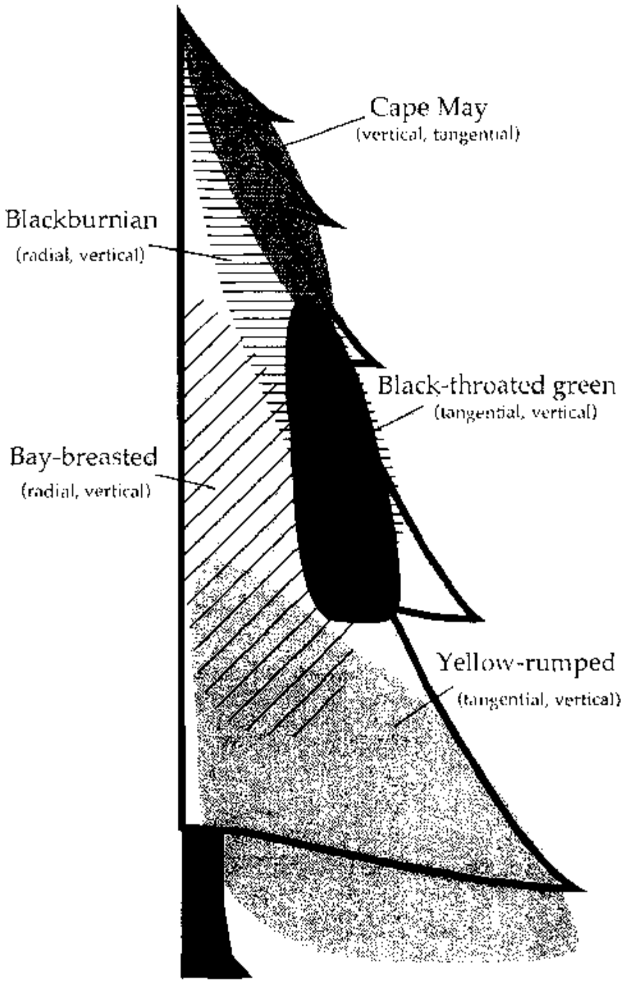


FIGURE 5 Niche relations among some North American warblers that forage in coniferous forests. Species differ according to microhabitat (parts of the tree most frequently visited) and foraging behavior. From MacArthur, 1958.

Niche theory is the branch of ecology that has taken these informal graphical models and extended them mathematically to model the process of competition and community organization. The origins of niche theory can be traced back to the early work of Volterra, and has grown and undergone dramatic changes. Complex interactions among members of a community are very difficult to measure, and the approach of studying theoretical models has greatly enhanced our knowledge of the factors that influence community dynamics.

B. Other Niche Dimensions

The classical context of niche comparisons is mainly restricted to niche axes characterizing habitat and food resources. Some ecologists have even considered the definition of a niche to be limited to the food resources consumed. However developments in the field of ecology through the 1900s have made it clear that an expanded definition of the niche that encompasses all the interactions of a species within a community is most appropriate. A broad definition of the niche includes all potential interactions that ultimately have the effect of changing the population density of another species in the community.

Mutualistic interactions are somewhat different from food resources in that they are often not literally consumed, nor are they occupied as in a nesting or sheltering place. However in niche descriptions, mutualistic interactions are treated much the same as other resources in that they may be limiting and therefore may even be competed for. For instance, pollinators interact mutualistically with angiosperms, yet from the point of view of a plant, they may be regarded as any other resource.

C. Indirect Interactions

Competition for resources has historically played a central role in niche descriptions. More recently, studies have uncovered processes where two or more species coexist because of factors that have little to do with resource niche axes. For example, predation can act to promote coexistence of potentially competing species in the following way: The presence of a predator may reduce the numbers of one prey species and thereby allow other species that have similar resource requirements to coexist. If two or more species are all harvested by a predator that does not show preferences for any specific prey type, then the most common prey will experience a disproportionate amount of predation. In this way, the presence of a predator can actually promote coexistence of species and an increase in biodiversity. Predators that have this effect have been referred to as keystone species because of their disproportional affect on community diversity.

An excellent example of this process can be found in the rocky intertidal communities of the eastern Pacific. Paine (1980) and colleagues conducted a unique set of experiments demonstrating that the presence of a starfish predator (*Pisaster*) enabled the coexistence of a number of species, whereas in the absence of the starfish, the community was dominated by the California mussel. The niche description for a barnacle in this community would therefore be incomplete without incorporating information about not only the presence of the California mussel, but also the relative vulnerability of each of these species to starfish predation. This implies that the niche characterization may change depending on the presence of potential predators and prey and their specific traits.

Predation certainly can be seen as a direct interaction, but the example above falls into the class of indirect interactions. A keystone predator can enhance the abundance of a species by interacting with a third species. Many interactions may cause ripple effects through a community in indirect ways. Elephants change the physical structure of their habitats, and many species excavate holes and nests that are used by other species. Parasites and pathogens can also have indirect effects on entire communities.

Viruses and other pathogens may initially appear to occupy a very simple niche, especially those that are confined to live entirely inside the body of a host species. How can such an organism play an important role in the community through indirect interactions? Pathogens can significantly impact the population size of the host species. This change in abundance may have consequences for other species, such as predators and prey of the infected species.

Some pathogens have complex life cycles that depend on more than one host species, and sometimes these hosts are very different. For instance, in using humans and snails at different stages of its life cycle, Schistosomiasis provides a clear link between humans and some snail species that are used as hosts for different stages of the life cycle.

Another interaction involving pathogens provides an example of apparent competition. One host species may be relatively unaffected by a pathogen, whereas other species may be severely impacted. In this case, the unaffected species can act as a carrier, spreading the pathogen to other species in the community that are more susceptible, causing their decline. On the surface, this pattern may resemble competition.

Community ecologists have used various methods to incorporate these indirect interactions into model representations of communities. It is certainly not as straightforward as simple resource utilization graphs. Mathematical matrices have been applied to study the dynamics of communities. In this format, all pairwise interactions can be explicitly considered, and broad concepts such as the relationship between community stability and complexity have been studied within this framework. Recently, web theory has been applied to this task because complex interactions within communities can be more explicitly constructed. All of these theoretical constructs of communities are based on the niche concept, thus the utility of the niche concept has grown even as the field of ecology has undergone dramatic changes.

IV. EVOLUTION OF THE NICHE

Ecological studies can help us to understand *why* species occupy unique niches, but *how* did the great diversity of species arise in the first place? The processes underlying the evolution, adaptation, and coexistence of species that have unique niches is a fundamental aspect of understanding biodiversity.

For many organisms the primary mode of species formation, and therefore the primary generating force for biodiversity, is thought to involve differentiation of isolated populations. A terrestrial species may have populations separated by water, or coastal marine populations may be isolated by unsuitable open ocean. Over time, these populations undergo changes and begin to accumulate different characteristics. Some of these differences are due to chance events such as mutation and genetic drift. Traits change over time as the organisms adapt to the unique features of their respective habitats. Subtle environmental differences may result in the evolution of thicker shells in a population of crabs, or a preference for fish over mammals in a population of killer whales.

This differentiation by adaptation to different habitats with different food resouces is in itself one engine for generating biodiversity, however it cannot explain the vast diversity of species that occur within a single habitat. Eventually, populations that have been isolated for a while may reestablish contact, and individuals will attempt to coexist within the same community. It is at this stage of speciation, referred to as secondary contact, that one can see how the concept of the niche is fundamentally related to species diversity.

When secondary contact occurs, one of three outcomes is possible. First, populations may interbreed, and speciation is not completed. Second, if populations do not interbreed, they may either coexist, or, third, one may outcompete the other. Which of these two latter outcomes is realized will depend on the amount of niche overlap between these species. If overlap is high, then one species is likely to outcompete the other. If overlap is low enough, the species may coexist.

In situations where coexistence is possible (or at least extinction is not very rapid), and the species involved still possess similar traits, we would expect evolutionary changes to occur over time. Individuals that consume limiting resources that are shared with another species will experience more competition and may survive or reproduce less than individuals that do not overlap as much. If there is natural variation in resource consumption among individuals in each population, and if the traits that cause this variation are genetically passed to offspring, then the species are expected to diverge over time to minimize niche overlap. This process is called ecological character displacement, and it is a logical extension of the competitive exclusion principle into evolutionary time. Figure 6 illustrates the process of character displacement. It is the change in a trait of a species, such as body size, that leads to a shift along a niche axis and reduction of overlap.

The empirical evidence for character displacement lies mostly in descriptive, snapshot-like comparisons of species. Typically, closely related communities are compared and character displacement is inferred if the presence or absence of one species is correlated with a change in the traits of another species. These studies infer that it is niche overlap in the use of a single limiting resource that has led to the observed species differences. Lack (1947) was one of the first to infer character displacement in this way in his study of Darwin's finches Figure 7.

Static character differences constitute indirect evidence for competitive processes and have therefore received criticism. To document evolution directly in natural populations is a difficult task because even rapid

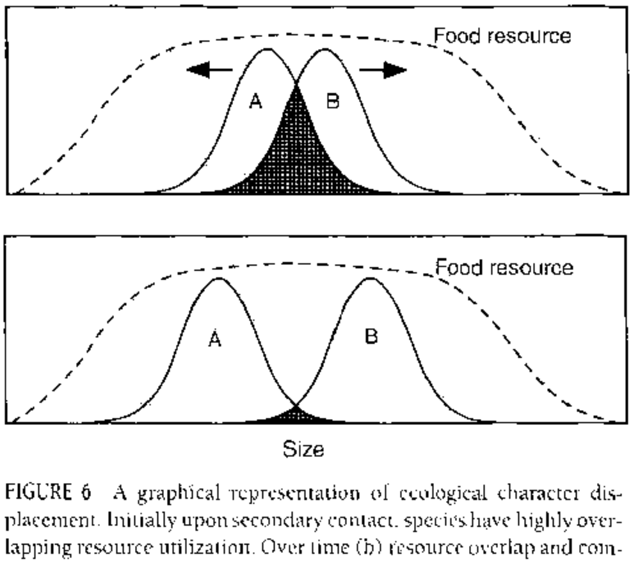


FIGURE 6 A graphical representation of ecological character displacement. Initially upon secondary contact, species have highly overlapping resource utilization. Over time (b) resource overlap and competition is reduced through evolutionary changes that result in a shifting of the niche.

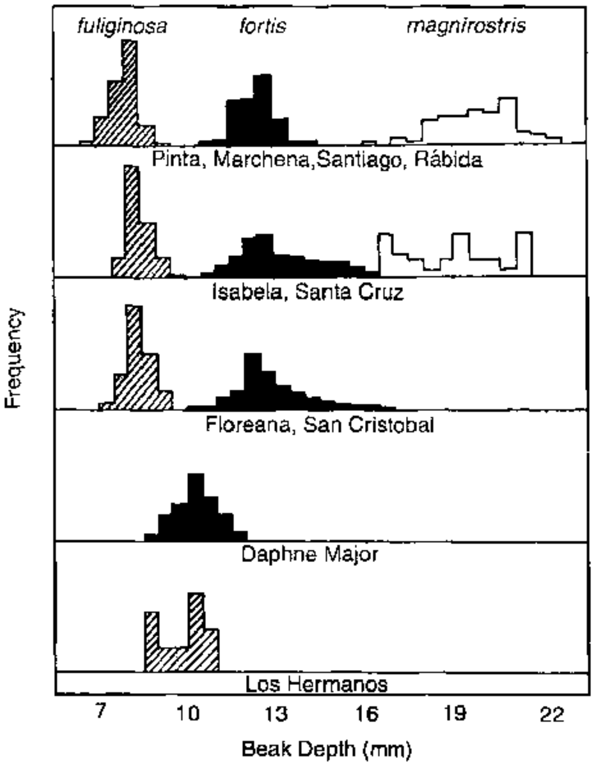


FIGURE 7 Size variation among three species of Darwin's finches (genus *Geospiza*) that specialize on seeds. Individuals are pooled among islands (given) to generate the histograms of beak depth. Populations on islands with fewer species (lower two panels) are intermediate in size compared to islands with the complete complement of species (upper panels). This is indirect evidence for ecological character displacement caused by competition for seeds; Pennsylvania in the United States. From Grant and Schluter, Ecological Communities. Copyright © 1984 by Princeton University Press. Reprinted by permission of Princeton University Press.

evolutionary changes may require generations to occur. Nevertheless, there is solid empirical evidence for niche evolution in nature. The Grants and their colleagues (see Grant, 1999) have shown the niche of Darwin's finches evolved over time according to the seeds that are available. In unique experimental tests of character displacement, Schluter (1994) has shown that stickleback fish populations changed over time in response to competition and niche overlap. Natural selection acted against individuals with feeding anatomies that caused them to harvest resources that were shared between species.

Ecologists may differ in opinion about the importance of character displacement in specific instances, and there are surely other very important forces that act to shape community composition and species traits. Nevertheless, few would doubt that character displacement is an important fundamental process in shaping communities of organisms.

V. THE HABITAT AND BIODIVERSITY

We have traced the habitat concept from its independent origin through its incorporation into the concept of the niche. With growing awareness of the need to conserve biodiversity, the habitat concept has emerged again as an important tool for management. Essentially, habitats contain functioning communities of organisms. If the goal is to protect a focal species, then setting aside appropriate habitat will likely result in preserving the required niche elements of the species as well.

Assessing appropriate habitat has two great advantages over attempting to define niches. First, because of the physical nature of habitats, they are easier to recognize and quantify. Second, ecological communities tend to fall along similar boundaries as habitats, thus by preserving a habitat one is more likely to preserve communities with all niche interconnections intact. This relationship among species, communities, and habitats has become the center of the more recent conservation efforts termed habitat conservation plans.

For example, in San Diego county, a number of endangered or threatened species rely on a unique type of habitat: chaparral, or coastal sage scrub. A major goal of habitat conservation plans in this region is to create contiguous regions of appropriate habitat. In this way, the needs of many species can be met at once.

Habitat fragmentation is a major concern because small refuges of habitat are essentially islands. Islands are prone to a number of processes that degrade biological diversity. Small populations on islands are prone to extinction through random fluctuations in numbers. The geometry of islands gives them a high ratio of edge to interior. Because animals may wander across this boundary, those that rely on the interior island habitat experience more competition and predation from those outside the island. More isolated islands are also less likely to be recolonized in the event that a population has gone extinct. Therefore, natural reserves and habitat conservation plans pay particularly close attention to fragmentation and try to consolidate larger parcels of unbroken natural habitat.

Once a habitat has been severely degraded, there may be ways to return it to its former natural state. Habitat restoration is a challenging field, but one that has made great strides recently. Wetlands have been restored by controlling of the water table and planting appropriate vegetation. Foresters have been restoring habitat for decades by replanting harvested areas. Highway construction projects aim to restore natural vegetation upon completion of a project by spraying a mix of grass, flower, and shrub seeds in a fertilizing matrix onto the bare ground. Another interesting example of habitat restoration can be seen with the construction of artificial reefs. By placing physical structures on the seafloor, a reef community can take hold and flourish. This is an example of how subtle environmental features such as physical structure can have a profound effect on habitat quality and species diversity.

Two general mechanisms act to promote species diversity within structurally diverse habitats. First, biological species diversity leads to more species diversity. A monoculture such as an orchard allows for relatively little niche differentiation of the organisms within. There is little opportunity for species to adapt to specific host species, and this applies to insects, birds, mammals, and parasites. Reduced numbers of species limits the number of possible niches in an orchard when compared to a tropical rain forest.

Artificial reefs belong to a second class of mechanisms that generate biodiversity by acting over and above the biological mechanisms. These are the purely physical features of the habitat that allow for increased niche differentiation. Dead logs, snags, distinct ground, and shrub layers will provide habitat heterogeneity that allow higher trophic levels to differentiate. Just as warblers partition the physical space within a single tree species, other competitors that use different tactics to forage on similar prey can partition a structurally diverse habitat to a finer scale. Behavior plays a key role as foraging tactics and escape strategies of prey are honed to specific microhabitats within the larger forest. MacArthur and colleagues (1962) were among the first to correlate purely physical aspects of a habitat with species abundances. Since then, physical habitat structure has been shown to augment species diversity in systems as diverse as birds, lizards, stream invertebrates, fish in marine reefs, and microorganisms in beakers.

Keystone species have a disproportional effect on many other species in the environment, and as such are often the focus of conservation efforts. Starfish foraging in the intertidal zone have a positive effect on biodiversity. Jaguars prey on a number of small rodents and animals, and they tend to capture prey in proportion to their abundance. The net effect is to prevent any single species from becoming very common and outcompeting other prey species. The niche of the jaguar and the starfish includes this community-wide role, and they are sometimes referred to as keystone predators.

Management strategies have turned more attention toward keystone species in order to preserve communities. A similar keystone role is played by tree species such as the Guanacaste tree of Central America. In this instance the link between the niche, the habitat and biodiversity is clear. The Guanacaste tree is vital for establishing forest in grassland habitats, and it therefore plays a key role in restoring tropical dry forest. The niche of this species would include its role in developing habitat structure and therefore intersects with the niches of many other species in the community.

The habitat and the niche can interact synergistically with human activity to cause extinctions and loss of biodiversity. Niche complementarity has led to a very general pattern whereby species with similar niches are spatially segregated. Increased human movements on a global scale has had a dramatic effect in bringing these species into contact by introducing exotic species into new geographic regions. If these alien species have high niche overlap with resident species, the resident species can be driven to extinction by a successful invader. Even though there are many cases where introduced species can not establish or outcompete residents, many threatened and endangered species are at risk because of invaders. This threat also extends to introduced predators. Resident species that have adapted to a niche in one community can be driven to extinction by a predator that has evolved in a different community.

VI. CONCLUSIONS AND NEW FRONTIERS

The concepts of the habitat and the niche have grown mainly through study of the organisms most familiar to us as humans. Most conceptual advances were derived in some way from the study of larger animals and plants. Even within these groups, there has been a disproportionate amount of study of birds. Yet there have been significant contributions from other sectors of the biological world. For instance, Gause's study of *Paramecium* lies at the heart of the niche concept. It is fitting then that more attention is returning toward understanding the niches of microorganisms.

Experimental evolution using bacteria is a growing field that holds great promise. The short generation times and relatively simple ecologies of bacteria enable evolutionary experiments on the scale of the community. This is one of the few instances where the interaction of ecological and evolutionary factors can be studied during the process of community formation.

There is an enormous number of microscopic, planktonic and meiofaunal (0.4–1 mm) organisms about which we know very little. These organisms are ubiquitous, speciose, and show an amazing amount of diversity in form. However there are significant obstacles encountered in studying these smaller organisms that even make it difficult to quantify the habitats in which they can be found. Practical problems arise while trying to observe them in nature, while on the other hand, many fail to survive the transfer to laboratory environments. Yet there are compelling reasons to learn more about their ecology.

The niche of a terrestrial vertebrate or marine invertebrate is often defined largely by how it feeds or defends itself. In turn, anatomical structures have evolved that reflect these niche differences. Similar examples can be found in the nearly invisible world of small creatures around us. For instance, ciliates and other protozoans possess anatomical structures that suggest a diversity of roles even greater than those observed in larger organisms, yet there has been comparatively little study of their niches in nature.

Some organisms stretch the concepts of the niche and habitat to their limit. Some plants display an extreme amount of what is referred to as phenotypic plasticity: different individuals develop extremely different body forms depending on the environmental circumstances. Some ciliates have the ability to radically change the form of feeding or defensive appendages even within the lifetime of an individual.

Many bacteria, and even some larger meiofauna like the bearlike tardigrades, can enter a sporelike state to withstand extreme environmental conditions. They may persist for long periods of time in this state and may be very difficult to detect. Much like the seed bank of desert plant communities, these alternative states pose difficult problems in understanding the dynamics of their communities.

The habitat and the niche have grown, developed, and been applied with great success. Observation and theory have played important roles throughout, and experimental investigation has undergone a resurgence. Yet it is intriguing to ponder that most of this development has included only a subset of the biological world. Only time will tell whether the concepts of the habitat and niche will persist as the complex interactions of these diverse communities of small creatures are revealed.

See Also the Following Articles

CONSERVATION BIOLOGY, DISCIPLINE OF • DIFFERENTIATION • DIVERSITY, COMMUNITY/REGIONAL LEVEL • ECOLOGY, CONCEPTS AND THEORIES IN • GUILDS • SPECIES DIVERSITY, OVERVIEW

Bibliography

Brown, J. H. (1975). Geographical ecology of desert rodents. In *Ecology and Evolution of Communities* (M. L. Cody and J. R. Diamond, Eds.), pp. 315–341. Harvard University Press, Cambridge.

Elton, C. (1927). *Animal Ecology*. Sidgwick & Jackson, London.

Gause, G. F. (1934). *The Struggle for Existence*. Williams & Wilkins, Baltimore.

Giller, P. S. (1984). *Community Structure and the Niche*. Chapman and Hall, London.

Grant, P. R. (1999). *Ecology and Evolution of Darwin's Finches*, second printing. Princeton University Press, Princeton.

Grant, P. R., and Schluter, D. (1984). Interspecific competition inferred from patterns of guild structure. In *Ecological Communities* (D. S. Strong, D. Simberloff, L. G. Abele, and A. B. Thistle, Eds.), pp. 201–233. Princeton University Press, Princeton.

Grinnell, J. (1917). The niche-relationships of the California thrasher. Auk, 34, 427–433.

Hutchinson, G. E. (1957). *A Treatise on Limnology. I. Geography, Physics and Chemistry*. Wiley, New York.

Hutchinson, G. E. (1959). Homage to Santa Rosalia, or why are there so many kinds of animals? *Am. Nat.* 93, 145–159.

Hutchinson, G. E. (1978). *An Introduction to Population Ecology*. Yale University Press, New Haven.

Lack, D. L. (1947). *Darwin's Finches*. Cambridge University Press, Cambridge.

MacArthur, R. H. (1958). Population ecology of some warblers of Northeastern coniferous forests. *Ecology* 34, 599–619.

MacArthur, R. H., MacArthur, J. W., and Preer, J. (1962). On bird species diversity II. Prediction of bird censuses from habitat measurements. *Am. Nat.* 96, 167–174.

Paine, R. T. (1980). Food webs: linkage, interaction strength and community infrastructure. *J. Anim. Ecol.* 49, 667–685.

Pennak, R. W. (1951). Comparative ecology of the interstitial fauna of fresh-water and marine beaches. *Ann. Biol. Sér.* 3, 462–463.

Root, R. B. (1967). The niche exploitation patterns of the blue-grey gnatcatcher. *Ecol. Monog.* 37, 317–350.

Schluter, D. (1994). Experimental evidence that competition promotes divergence in adaptive radiation. *Science* 266, 798–801.

Terborgh, J. and Weske, J. S. (1975). The role of competition in the distribution of Andean birds. *Ecology* 56, 562–576.

Volterra, V. (1926). Variations and fluctuations of the number of individuals in animal species living together. Reprinted in 1931 in *Animal Ecology* (R. N. Chapman.) McGraw-Hill, New York.

Whittaker, R. H., and Levin, S. A. (1975). *Niche: Theory and Application*. Dowden, Hutchinson & Ross, Stroudsburg.

HEMIPARASITISM

David Smith,* Todd J. Barkman,† and Claude W. dePamphilis†
*University of Maryland and †Penn State University

GLOSSARY

dodder Viney parasites of the genus *Cuscuta* that form haustoria on the stems of their hosts. The unrelated parasitic vine, *Cassytha*, is often referred to as Laurel Dodder.

endophyte The portion of a parasitic plant that is embedded inside host tissue.

endophytic parasite Parasites with vegetative bodies that are entirely endophytic and unobservable unless flowering.

haustorium (pl. haustoria) The organ found in all parasitic plants that penetrates vascular tissue of the host plant and forms a functional bridge for uptake of water and nutrients from the host.

hemiparasite A parasite that photosynthesizes.

holoparasite A parasitic plant that lacks chlorophyll and cannot photosynthesize.

mistletoe Parasites of the closely related families Loranthaceae or Viscaceae with haustoria that originate as root tissue and penetrate stems or roots (in Loranthaceae) of the host plants.

mycotrophic (=mycoparasitic) plant A plant that obtains energy through mycorrhizal associations with saprophytic or other biotrophic fungi.

root parasite Parasites that form haustoria on roots of the host plant.

stem parasite Parasitic plants that form haustoria on stems of the host.

AROUND 3000 SPECIES OF plants are considered parasites because they obtain water, mineral nutrients, sugars, and sometimes other materials from another plant. Parasitism in plants has much in common with other plant life-history strategies, but it is distinguished by having a direct, physiological connection called a haustorium. Modern techniques for phylogenetic reconstruction have demonstrated that parasitism has originated a dozen times within the flowering plants, including mistletoes, dodders, and many agriculturally important species. The ecology of the parasitic plants and the interaction with the host are briefly described.

I. THE NATURAL HISTORY OF PARASITIC PLANTS

Parasitic plants are strictly defined by the production of specialized feeding structures, known as *haustoria*, that form a functional bridge into their hosts. Through the haustorium, parasitic plants obtain water, mineral nutrients, sugars, and other host metabolic products. Approximately 3000 angiosperm species (1% of named species) are haustorial parasites; however, this total does not include any plants that feed heterotrophically on fungi (*mycotrophic* plants). Most parasitic plant species are *hemiparasites*; they photosynthesize and are only partially heterotrophic. The degree of heterotrophy in parasitic plants varies between autotrophy and the complete loss of photosynthesis, a condition known as *holoparasitism*.

Other than the presence of haustoria, there are few generalizations that can be made about parasitic plants because they display enormous variability in growth form and host dependence. In most species, haustoria are formed from root tissue of the parasite, but in a few species they are initiated from stem tissue. The haustorium, in turn, can penetrate the root or stem of the host plant. The major life-histories among parasitic plants are distinguished by these differences and they have important ecological implications.

Root parasites are extremely variable in growth form, occurring as trees, shrubs, and herbaceous perennials or annuals. Because their haustoria are often below ground, many root parasites do not demonstrate any obvious indications of parasitism. Experimental studies have compared hemiparasitic plants grown in pots with and without hosts. They found enormous variability among species in their ability to set seed in the absence of a host. The degree of parasitism is extremely variable in root parasites and the increase in seed set for any one species is different depending on the host.

Dodders and some *mistletoes* are both called *stem parasites* because they attach to the stems of the host; however, they have very different life histories. Mistletoes are closely related to root parasites and, in fact, their haustoria are derived from root tissue. Upon seed germination, their haustoria immediately penetrate the host's stem. Mistletoes most often grow on one individual, but in some cases they may sprawl along or hang down from the stems of the host plant. In contrast, dodders are closely related to nonparasitic vines. After germination, dodder seedlings begin to sprawl and twine along the stem of the host, forming haustoria from time to time where the new stems of the dodder contact the host. The rudimentary hypocotyl of the dodder eventually dies and it grows extensively beyond the original point of germination, often using multiple hosts.

All parasitic plants are obligate, requiring a host for survival, growth, or seed production, except for some root hemiparasites. The degree of parasitism is a useful concept for comparing the ecology of related species that vary in their dependence on the host. At one extreme, the ecology of facultative parasites is similar to their fully autotrophic relatives. At the other extreme are a few highly derived lineages of nonphotosynthetic *endophytic parasites* that have all of their plant body within the host tissue emerging only to flower and fruit. Among root holoparasites, there are enormous differences in the amount of biomass that is below ground. Some species are almost exclusively subterranean, growing as long-lived underground tubers, producing only flowers and fruits above ground.

Extremely derived holoparasitic plants are some of the great oddities of nature. In the 10th century, Arabic scholars wrote that parasitic plants performed the actions of animals by feeding from the juices of the host plant; they have the "souls of worms." In 1828, a naturalist named Trattanick proposed that a separate category should be created to encompass them, the taxonomic equivalent of a mental asylum for the manic and delusional. Certainly the morphology of many holoparasites is nonplant like, and this led some early researchers to call them fungi!

The dissonance experienced on first learning about parasitism in plants betrays a common prejudice in some botanists; plants and parasites are naively regarded as opposites. As a result, parasitic plants have been misidentified and misunderstood. The "bearded-grape" was regarded as a monstrosity and assigned to its own genus; however, it was simply a grape infected by *Cuscuta!* Parasitism in most root hemiparasites was unnoticed by naturalists until very recently making the modern view of parasitic plants only about a hundred years old. In 1969, Job Kuijt's book, *The Biology of Parasitic Flowering Plants*, abolished many of the mysteries surrounding parasites. In this seminal work, Kuijt provided a critical analysis of all previous research and proposed a conceptual framework for understanding the evolution, morphology, physiology, and ecology of parasitic plants. While Kuijt's book remains the best authoritative source on the subject, new molecular evidence is being used to resolve the detailed phylogenetic history of parasitism, and a wide range of biochemical, genetic, and molecular approaches are now being applied to better understand parasite biology.

II. THE PARASITIC PLANT CLADES AND THEIR RELATIONSHIPS TO OTHER PLANTS

Phylogenetic relationships between parasitic plant families and their affinities to nonparasitic plants are now well known for most species. A few families of root holoparasites have been difficult to place in a phylogeny because their floral morphologies are unique and they virtually have no vegetative morphology. Molecular phylogenetic study of many parasitic plant groups has been impaired by the fact that photosynthetic and other genes commonly used in plant molecular systematics are missing or evolve at greatly accelerated rate. However, recent studies of plant mitochondrial DNA sequences have resolved 11 or 12 independent origins of haustorial parasitism (Fig. 1) distributed throughout angiosperm history.

The Sandalwood order (Santalales) are the largest, most diverse clade of parasitic plants in terms of growth habit, but all retain at least some chlorophyll and photosynthetic ability; in fact, some members may not even be parasitic. This group includes almost 2000 species of trees, shrubs, mistletoes, and at least one species of mainly endophytic mistletoe. Olacaceae and Opiliaceae are composed entirely of trees and shrubs that include both free-living as well as root parasitic members. Santalaceae, are also trees and shrubs with both root and stem parasites. The mistletoes belong to one of two large families, Loranthaceae and Viscaceae. The stem and root parasites of Loranthaceae are known as showy mistletoes because of their long, colorful, tubular flowers. Viscaceae, including the "dwarf mistletoe," *Arceuthobium*, are all stem parasites. In addition to the two mistletoe groups, two other families of Sandalwoods, Misodendraceae and Eremolepidaceae, are stem parasites.

About 1700 root parasitic annuals and perennials are found in what has traditionally been classified as two closely related families, Scrophulariaceae and Orobanchaceae. As a whole, they represent the most diverse clade of parasitic plants in terms of variation in degree of heterotrophy. Traditional Orobanchaceae includes only holoparasites, while the large family Scrophulariaceae (ca. 4500 spp.) has nonparasites, hemiparasites and even some holoparasites. Molecular phylogenies based on chloroplast gene sequences have found the distinction between the two families to be artificial—a strongly supported clade contains all known parasites from both groups, but most of the rest of traditional Scrophulariaceae are not closely related to this group. In other words, Scrophulariaceae appear to be polyphyletic, while the parasitic lineages are quite clearly monophyletic, suggesting a single origin for parasitism in this group. For this reason, dePamphilis and colleagues have proposed to include all of the parasites (plus a few closely related nonparasitic genera) in an expanded Orobanchaceae. Within this large group, holoparasitism appears to have evolved not less than five times. The agriculturally important genera *Striga* and *Alectra* (both mostly hemiparasitic) and *Orobanche* (holoparasitic) are members of this clade, as are the large hemiparasitic genera *Pedicularis, Euphrasia*, and *Castilleja*.

The remaining parasitic plant families are not nearly as large or diverse as Orobanchaceae or Santalales. Plants commonly referred to as dodders are found in two families Cuscutaceae and Lauraceae. Cuscutaceae is a family of about 150 species of one genus, *Cuscuta*, that are distributed nearly worldwide. Lauraceae is a large family comprised of mostly nonparasitic trees that also includes one hemiparasitic genus, the vine *Cassytha*. *Cassytha* has about 20 species, most of which are endemic to Australia or Africa, with one species found throughout the tropics. Although strikingly similar in gross morphology, these two parasitic groups originated independently. *Cuscuta* is related to morning glories (Convolvulaceae), in the order Solanales, while *Cassytha* belongs in Laurales, a basal angiosperm order. *Cuscuta* are usually considered to be holoparasites, but very low photosynthetic activity exists in some species; whether or not the chlorophyll-containing species can use their photosynthetic products is unknown.

Approximately 50 species of endophytic holoparasites are found in the traditionally recognized family, Rafflesiaceae. Although considered to be closely related, molecular phylogenies now indicate up to four independent origins (polyphyly) for the genera making up this collection of extreme parasites. Molecular evidence is in agreement with previous studies of pollen morphology that suggested Rafflesiaceae be divided into four families: Rafflesiaceae, Mitrastemonaceae, Apodanthaceae, and Cytinaceae. While floral morphology is diverse for these families, they were previously thought to be related because they all are endoparasites.

Rafflesiaceae includes the largest flower in the world, *Rafflesia arnoldii*, which measures 1 m across. *Rafflesia* species only parasitize species of *Tetrastigma* vines (Vitaceae) and have an unusual reproductive biology. Emerging buds develop for approximately 1 year before they open and become receptive to pollinating flies that are attracted to the often fetid scent. Flowers die within 5 days leaving a narrow window of time for pollination to occur. Because of the short flower life span, and

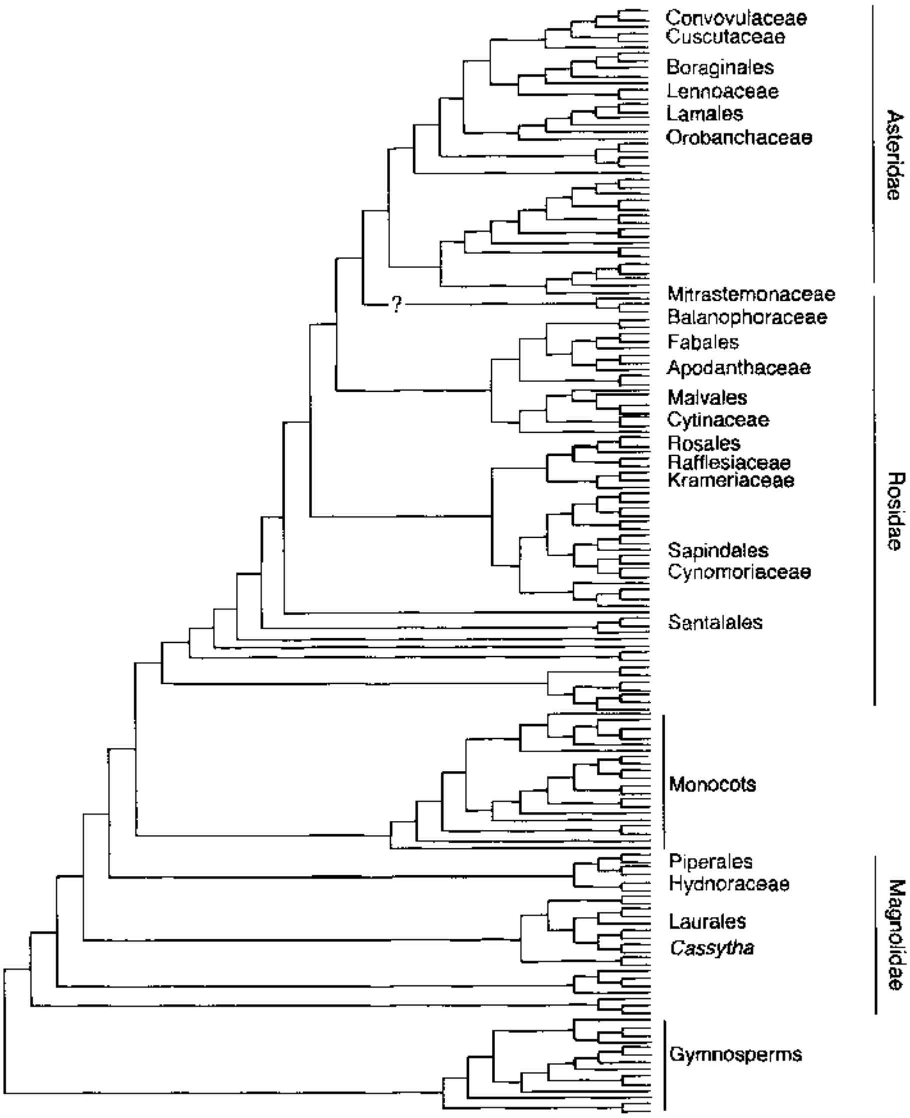


FIGURE 1 *Phylogenetic origins of parasitism in the flowering plants.* Summary tree of flowering plant phylogeny is based on the Neighbor-Joining analysis of mitochondrial gene sequences (*cox*1 and *atp*A) conserved in all parasitic plants regardless of photosynthetic ability (T. J. Barkman and C. W. dePamphilis, unpublished). Note that 12 or 13 independent origins are resolved, with parasitic groups containing holoparasites or only hemiparasites shown near their closest resolved nonparasitic group. Mitrastemonaceae + Balanophoraceae is annotated as uncertain (?) because monophyly of this group depends on choice of gene and method of phylogenetic analysis.

because female and male flowers are on separate plants, sexual reproduction is rare. Rafflesiaceae species are all exceptionally rare and endangered throughout their range, partly due to their reproductive strategy and partly due to habitat loss. As shown in Figure 1, recent mitochondrial DNA sequence data suggest that the closest relatives of Rafflesiaceae include the rose order (Rosales).

In contrast to *Rafflesia*, members of Apodanthaceae, *Pilostyles* and *Apodanthes*, produce tiny flowers only a few millimeters wide. Approximately 20 species are included in Apodanthaceae, which are distributed from the southwestern United States throughout the northwestern tropics with restricted localities in Africa, Australia, and the Middle East. Curiously, Apodanthaceae appear to be most closely related to the order Fabales, which includes their primary hosts: various species of legumes.

Cytinaceae include two genera, *Cytinus* and *Bdallophyton*. *Cytinus*, with 6 to 7 species, is distributed in the Mediterranean region, South Africa, and Madagascar and most commonly parasitizes species of *Cistus* (Cistaceae). *Bdallophyton* has only 1 to 4 species that are found mainly in Mexico and parts of Central America usually parasitizing members of the Burseraceae. The closest relatives of this family are classified within the cotton or hibiscus order, Malvales. Like the previously mentioned families, Rafflesiaceae and Apodanthaceae, Cytinaceae are not known to cause economic loss because they do not parasitize crop plants.

Mitrastemonaceae are specialists on trees in the family Fagaceae found in southeast Asia north to Taiwan and parts of central and South America. *Mitrastemon* is the only genus included in this family that is made up of two named species that are totally white. Plants are protandrous and may produce large quantities of nectar.

Another small family, Hydnoraceae, is composed of about 20 species of root holoparasites native to tropical and southern Africa, Madagascar, and South America. Hydnoraceae have traditionally been considered close relatives of Rafflesiaceae, based largely on the shared holoparasitic habit with a reduced vegetative body and highly modified floral structures. However, molecular phylogenies now suggest that Hydnoraceae represents an independent origin of parasitism within the basal angiosperms, and that similarities with other holoparasitic groups have evolved convergently.

Balanophoraceae has about a hundred species in 17 genera, all of which are extremely reduced root holoparasites that have some of the smallest flowers in the world. Members of Balanophoraceae are found throughout the tropics and have broad host ranges but do not cause extensive crop damage. Balanophoraceae are sometimes included with the sandalwoods; however, molecular evidence does not support that relationship.

Cynomorium (Cynomoriaceae) is a holoparasite in the Mediterranean region comprised of two species that has sometimes been included within Balanophoraceae. As shown in Figure 1, Cynomoriaceae is not closely related to Balanophoraceae; instead it appears to be related to the order Sapindales, which includes the citrus family.

Krameriaceae includes 17 species of root hemiparasitic shrubs and herbs in the single genus *Krameria* that are often found in dry, open habitats in New World deserts and subtropics. *Krameria* species are unusual in that they produce floral oils that are used by oil-collecting bees. Traditionally, *Krameria* has been placed near the legumes (Fabaceae) or milkworts (Polygalaceae) due to a superficial resemblance of their flowers. Molecular evidence, however, unambiguously suggests a close relationship of Krameriaceae with Zygophyllaceae in spite of little morphological similarity between the two families.

Lennoaceae has two genera with five species of root holoparasites restricted to New World deserts in the southwestern United States, northwestern Mexico, and Columbia. All are herbaceous holoparasites of woody plants, including Boraginaceae (their closest known relatives) and Asteraceae. These strange parasites are spread by rhizomes often far beneath the sand or soil surface and are difficult to find unless flowering. *Pholisma sonorae*, also known as sand food, was traditionally eaten by indigenous people.

Finally, one gymnosperm genus, *Parasitaxus* (Podocarpaceae), is parasitic on other free-living Podocarpaceae in New Caledonia.

III. EVOLUTION OF THE DEGREE OF PARASITISM

The evolution of parasitism has been difficult to study because intermediate evolutionary stages between nonparasitic ancestor and extremely derived holoparasites are often missing. The variability in the degree of parasitism and the multiple origins of parasitism make these plants ideal for studying the evolutionary origins of parasitism and the often rapid canalization of parasitic traits. The degree of parasitism refers to increases in the dependence that parasitic plants have on their hosts. Aspects of the degree of parasitism include the amount of carbon, water, or sugar obtained from the host, as well as major changes in the way that parasitic plants relate to a host. These major shifts include the evolution of stem parasitism or the evolution of an endophytic life history. The evolutionary factors that have influenced these events remain one of the most important questions about parasitic plants.

Changes in the degree of hetrotrophy represents one of the most interesting and best-documented cases of quantitative evolution of parasitism. Some hemiparasitic root parasites and mistletoes are almost entirely autotrophic—that is, they obtain a small fraction of their carbohydrates from their host whereas holoparasitism (100% heterotrophic) occupies a presumably irreversible endpoint of this quantitative scale. Most clades of parasitic plants are entirely hemiparasitic or entirely holoparasitic. Krameriaceae are all hemiparasites, while the

Hydnoraceae, Rafflesiaceae, Apodanthaceae, Mitrastemonaceae, Cytinaceae, Lennoaceae, Cynomoriaceae, and Balanophoraceae are entirely holoparasitic. *Cassytha* and *Cuscuta* are often called holoparasites, although these species have very low amounts of chlorophyll. *Cassytha* species are yellow-green in color, while *Cuscuta* species are more orange-yellow. In *Cuscuta*, chloroplasts are less than 10% as numerous as in autotrophic, plants but the green color is masked by yellow pigments.

Members of the order Santalales are mostly hemiparasitic, but at least one species represents a very advanced stage of parasitism. In *Arceuthobium*, the endophyte is extremely well developed and the leaves are reduced to scales. The most derived parasite in the Sandalwoods is found in the endophytic parasite, *Tristerix aphyllus*. *Tristerix aphyllus* is a parasite of columnar desert cacti, whose endophytic life history may allow the parasite to escape the hot and desiccating desert conditions. Curiously, *T. aphyllus* and all *Arceuthobium* species retain some chlorophyll, and in spite of these extreme advances toward parasitism, true holoparasites are absent in the order.

In contrast to the conditions in Santalales, within Orobanchaceae, holoparasitism has evolved on several occasions. In fact, some species include hemiparasitic and holoparasitic populations and the degree of heterotrophy varies substantially. Peter Atsatt (1970) was able to produce holoparasites in a few generations in artificial selection experiments. This suggests that the complete evolutionary loss of autotrophy, as seen repeatedly in Orobanchaceae, may be relatively common in nature. One obvious conclusion is that holoparasitism is more likely to evolve in root parasites than in stem parasites. Since root parasites are often found in the dark understory, the selective advantage of retaining functional chloroplasts in root parasites may be much weaker than for sunbathed stem parasites. In fact, holoparasitism in beech drops (*Epifagus virginiana*, Orobanchaceae) is clearly irreversible because entire sections of the genome that contain the genetic instructions for photosynthesis have been deleted.

The evolution of the mistletoe life history represents a different kind of innovation in the degree of parasitism that is different from the evolutionary loss of chlorophyll. Stem parasitism may have evolved more than once within the Sandalwoods; two families, Santalaceae and Loranthaceae, have both stem and root parasites. It is possible that these represent paraphyletic taxa, but it is also possible that study of intermediates between root and stem parasites in Santalales will lead to an understanding of the evolutionary forces that led to the evolution of stem parasitism from root parasitism.

IV. PARASITISM AND OTHER PHENOMENA IN PLANTS

The utility of the strict botanical definition of parasitism is most apparent when parasitic-host interactions are compared with other plant-plant interactions. Almost all plant-plant interactions are characterized by conflict, and many are ecologically parasitic because one species exploits another. In short, each aspect of parasitism is present in another kind of plant-plant interaction.

Parasitic plants acquire resources directly from another plant. Similar phenomena exist in root-grafted forest trees and in mychorrizal associations with fungi that indirectly connect plants. Root grafting is defined by a morphological continuum of the vascular tissue between two individuals that occurs in several species of forest trees. Interestingly, enough materials can be transferred through natural root grafts to sustain girdled trees. While root grafting is most common between closely related species, parasitic plants often form haustoria on a variety of species that are not closely related.

Mycorrhizal associations are very common in nature. Some mycorrhizal plants, such as indian pipe (*Monotropa*, Monotropaceae) have evolved complete heterotrophy and lost functional chloroplasts, much like holoparasites; actually these myco-heterotrophs are often confused for parasitic plants. In general, mycorrhizal associations are not species specific, and networks of mycelia and roots from a number of individuals can be interconnected thereby allowing the exchange of resources. In root grafts as well as mycorrhizal associations, source-sink relationships may exist between individuals; however, the precise relationship depends on ecological differences that may change over time. In haustorial parasitism, the parasitic plant is generally the sink while the host is the source.

Sometimes plants exploit the habitat created by another plant at the expense of the "host." For example, epiphytes use hosts primarily as anchors. Epiphytes intercept light, minerals, and water that might otherwise be available to the host plant. The reduction of available resources, plus the mechanical stresses on the supporting structure, may impose fitness costs on the host plant. A phenomenon related to epiphytism, but potentially more harmful to the "host," involves nurse-plant relationships, in which an established plant creates a microhabitat that facilitates the establishment of another species. For example, vines use the structural wood of trees to reach the forest canopy without investing in support themselves. In an extreme case, strangler figs twine around the trunk of a host, sending down adventitious roots to eventually support itself. Often the fig grows into a large tree that densely shades and envelops the host and may kill it. In contrast, parasitic plants impose fitness costs by removing water, mineral, and carbon resources from the host plant. They may also physically damage conducting tissues and negatively affect the host's metabolism through hormonal interactions.

V. SEED DISPERSAL, GERMINATION, AND PREHAUSTORIAL EVENTS

Before forming haustoria, the seeds of a parasitic plant must germinate close to a host plant and make contact. Seeds have little control over where they are dispersed and because seedlings are fragile, encountering a host is an important but difficult challenge.

New contacts between hosts and root parasites are initiated by root growth; either the host grows near the parasite or the parasite grows toward the host. Some species use chemicals exuded by host roots as a cue for finding the host such that seedlings grow chemotropically up gradients. In facultative root parasites, the plants can invest in root systems that "search" for hosts. In the meantime, these roots can absorb resources and the plants can usually persist until they contact a host. In contrast, obligate parasites do not invest in elaborate structures; they must make contact immediately or die.

Seed size and number and the degree of parasitism are correlated in parasitic plants. Holoparasites generally produce large numbers of small seeds. This correlation is probably driven by different strategies for managing the risk of finding a host. Producing larger, more provisioned seeds allows the facultative parasite to persist until it finds a host, but the production of many small seeds increases the probability that at least some seeds will land near a susceptible host root. In general there is a negative correlation in all plants between seed size and number, which is thought to be related to resource limitations. Theoretically, resources are unlimited for parasites unless the host is killed. Therefore resource limitations may not be the reason for the seed size-number tradeoff in parasites.

In some species of obligate root parasites, seed dormancy may be prolonged as a way to spread the risks associated with finding a host over time as well as space. In other words, if a seed is dispersed to a suboptimal site, dormancy allows germination to occur at later times if the site becomes appropriate. Seed dormancy is broken when chemical cues indicate the proximity of a host. The first root exudate discovered to induce parasite seed germination was Strigol. This chemical is found in cotton and it stimulates the germination of *Striga* seeds. It is not known how well seeds discriminate between chemical cues of hosts from nonhosts in their local environment, but seeds often make mistakes. For example, cotton has been used as a way of managing *Striga* in agriculture by causing seed germination in the absence of a suitable host (see Section X).

Mistletoe stem parasites must germinate on the stems of their hosts, requiring that seeds be dispersed directly onto a stem or leaf. Seeds are dispersed in two ways. Ballistic dispersal projects seeds up to 50 ft from the original host at velocities up to 90 ft/s as reported in *Arceuthobium*. Birds also disperse seeds of mistletoes over long distances because fruits are the main source of energy and protein for some birds. After consumption, seeds may be defecated onto hosts or the sticky seeds may attach to the bird's beak and be carried directly to another host. The seed may be wiped onto the host plant by the bird as it tries to remove the seeds from its beak.

VI. THE HAUSTORIUM

Haustoria are extremely varied structures among parasitic plants. In root parasites, haustoria are easily identified on exposed roots. They appear as swollen tissue at a contact point between parasite and host. In dodders, haustoria superficially resemble pegs or suction cups that connect the host and parasite. In general, haustorial cells occupy intercellular spaces and displace the host tissue, but enzymes also digest the host cell walls. Once a parasitic plant has encountered a host, it must penetrate the cambium and establish an interface. The interface is extremely varied. In most cases, the parasite forms a continuum with the xylem of the host plant, but in others (e.g., *Cuscuta*), the parasitic plant taps into the phloem. Host-derived materials may be transferred through strawlike intrusions into the host vascular tissue, or they are simply absorbed across cell walls. Haustorial cells near host tissue are usually rich in mitochondria and rough endoplasmic reticulum suggesting they are probably actively producing proteins that are likely used to produce digestive enzymes. After forming the initial haustorium, parasites may enhance local root growth to increase the number of haustoria and strength of the connection with the host. In stem parasitic mistletoes, seeds germinate and send out a radicle that grows into the host through a stomate.

VII. THE PHYSIOLOGY OF PARASITE—HOST INTERACTION

Once a haustorium has established, the question remains: Why do materials move from host to parasite? Part of the answer appears to be quite simple for most xylem-tapping species: parasitic plants usually transpire at higher rates than their hosts to maintain a gradient in water potential across the haustorial boundary. The stomata of most plants close at night to conserve water, and they may also constrict during the day to reduce rates of transpiration under drought conditions. However, in many parasitic plants, stomata remain open at night, and they tend to maintain a high of transpiration throughout the day, even under drought conditions. Parasitic plants are among the least efficient plants at water use. In extreme cases, some species have glands to exude extra water. The net effect is that water and other dissolved substances move down the water potential gradient via mass flow from the host to the drier parasite.

If all the transfer of resources was facilitated in this manner, the concentration of nutrients should be higher in the parasite, but in the same proportion as the host plant. These general patterns are not found—instead, calcium generally has the same concentration as the host, while nitrogen, phosphorous, and potassium are enriched. Some evidence suggests that there is a component of active transport in parasitic plants. The haustorial cells at the interface have enhanced concentrations of mitochondria and show signs of being able to mobilize energy. It is possible that the main function of these cells is to pump unwanted materials back to the host. The potential benefits of reducing calcium concentrations and the mechanism(s) that make it possible are not fully understood.

VIII. THE ECOLOGICAL ADVANTAGES OF PARASITISM

In facultative parasites, the advantages of parasitism can be demonstrated experimentally. When grown with a host, parasitic plants look healthier, grow bigger, have more flowering branches, and produce more seeds. Compared with a patch of soil, the resources available in the vascular tissue of a host are extremely enriched. Parasitism allows plants access to a rich, hydroponic nutrient source. The benefit of parasitism can be mimicked experimentally by growing parasites in nutrient-rich solution culture.

A genus of annual hemiparasites, *Cordylanthus*, blooms during the hottest time of the year in an arid environment. Few other annual plants bloom at that time of year, and it is possible because the parasites use roots of perennial plants to gain access to groundwater. This unusual flowering time may also permit the parasites to have more exclusive access to pollinators.

In addition to nutrients and water, parasitic plants absorb secondary chemicals from hosts. Recent experiments by Michelle Marvier showed that herbivory is slowed and herbivores have lower fitness on parasitic plants that are attached to hosts that deter herbivory. Parasitic plants use a wide range of hosts, and, as a consequence, may accumulate a diverse and changing complement of protective secondary compounds. This strategy may effectively make parasites "moving targets" that are difficult for herbivores to "hit" in ecological or evolutionary time.

IX. HOST SPECIFICITY

In general, parasitic plants are considered to be host generalists. The advantage of this is quite simple; seeds can not choose where they are dispersed. Typically, a higher fraction of mistletoe seeds successfully establish on the species that is most abundant locally. In addition, dodders preferentially coil toward good hosts. Measuring host selectivity is much more difficult in root parasites than stem parasites. However, careful measurements by Gibson and Watkinson have demonstrated that root parasites are generally selective.

Some species are only found on one species or genus. The entire family, Rafflesiaceae, is composed of host specialists, as discussed earlier. Two closely related species of Orobanchaceae, *Epifagus virginiana* and *Conopholis americana*, are specialized on beech and oak trees, respectively. *Striga* (Orobanchaceae) tends to utilize grasses, and some species are highly host specific, even to the extent of requiring a particular subspecies as a host.

A tempting generalization is that species that are more dependent on their hosts tend to be more host specific. Kuijt rejected this hypothesis because most species of root holoparasites are host generalists. However, the most host specific species, including *Epifagus* and *Conopholis*, Rafflesiaceae, and *Tristerix aphyllus*, all represent extremely derived parasites. It may be that the endophytic lifestyle or the subterranean habit releases these plants from other selective pressures, and not that they become host specific because they are extremely derived. New phylogenetic evidence may provide more rigorous tests of this hypothesis.

X. PARASITIC PLANTS AS PESTS

Many species of parasitic plants are economically important because they reduce crop yields. Parasitic plants remove nitrogen, phosphorous, and other mineral nutrients that may reduce the host's ability to grow. Parasitic plants have nutritional requirements that are very similar to their hosts, therefore the damage caused by a parasitic plant is often directly proportional to its biomass. In addition, parasitic plants are less efficient at water use than their hosts, so parasitized hosts will tend to be more water stressed than unparasitized hosts. Dodders do not have the same dramatic effects on water balance as mistletoes or root parasites, but their haustoria remove sugars and stunt developing fruits. In general, crops that are infested with parasitic plants grow slower, have lower yields, and are more susceptible to disease. In addition to crop plants, parasites are important pests of trees. The invasion of mistletoe haustoria into host trees can distort wood or alter growth patterns, thereby decreasing the value of the timber. Some mistletoes and dodders damage fruit trees or ornamental plants. The damage caused by these parasites is usually measured in lower yields, the loss of aesthetic value to ornamentals, or the increased effort required to control them. An enormous amount of research has focused on the root parasitic weeds, mistletoes, and dodders.

A. Root Parasitic Weeds

The most economically important root parasites are the witchweeds (*Striga*, Orobanchaceae). *Striga* includes approximately 17 species, of which, 11 are known to be pests. Most witchweeds have a tendency to parasitize grasses including sorghum, pearl millet, finger millet, rice, maize, and sugarcane. The most important witchweed species is *S. hermonthica*, which ranges from Ethiopia and Sudan through Sahel, north into Arabia, and south to the Ivory Coast, Nigeria, Angola, Namibia, and the Lake Victoria basin. *Striga asiatica* has a much wider range that includes much of Africa, parts of India, China, and Australia. In addition, *Striga asiatica* was accidentally introduced into North America in the 1950s. One species, *S. gesneroides*, has very different host range from the other witchweeds; it uses a diverse set of broad-leaved hosts. It is best known for the damage it causes to cowpeas in West Africa, but it also damages tobacco in East Africa.

Among root parasites, the holoparasitic broomrapes (genus *Orobanche*, Orobanchaceae) are second to the witchweeds in terms of their economic impact. The geographical ranges of the economically important species are centered in the Middle East, but they are found further west and south into Africa, west and north into Europe, east into India, Pakistan, and Nepal, and north into Afghanistan, and several countries that were once in the Soviet Union. In general, broomrapes have wide host ranges. *Orobanche ramosa* attacks crops in Solanaceae, Brassicaceae, and Fabaceae. It sometimes attacks onions, but in general it is never found on grains (it was once reported on maize). The host range of *O. cernua* is more restricted; it parasitizes only Asteraceae (mainly sunflower) and Solanaceae (tobacco, tomato, and eggplant). *Orobanche crenata* Forsk. has a wide host range, but it is most important as a pest of Fabaceae, especially faba beans and some Apiaceae (carrots).

The genus *Alectra* (Orobanchaceae) is closely related to the witchweeds. Of the 30 species, 4 are notable pests. *Alectra vogelii* and *A. picta* are pests of cowpeas and other pulse crops in semiarid Africa. *Alectra orobanchoides* occasionally attacks sunflowers or tobacco in South Africa and *A. fluminensis* parasitizes sugarcane in Central and South America. Several other genera in Orobanchaceae have agriculturally important species, but they are not nearly as important as the witchweeds, broomrapes, or *Alectra*. Other genera with notable or potentially important pest species include *Buchnera, Ramphicarpa, Odontites, Rhinanthus, Aeginetia, Melampyrum*, and *Christonia*.

Most control methods for root parasites focus on reducing host damage as well as seed set. Witchweeds and broomrapes produce hundreds of thousands of small, dust-like seeds per plant. Furthermore, the seeds can lie dormant in the soil for several years, and therefore reducing the size of the seed bank involves enormous effort. The most effective available method of control involves hand pulling the plants before they set seed. Although most of the damage to the host has been done by the time the parasitic plants emerge, hand pulling can reduce some of the damage. More important, early hand pulling prevents reseeding. If reproductively mature plants are pulled, care should be taken to prevent the seeds from being dispersed, and the adult plants should be burned.

More direct chemical methods can be used to reduce the seed bank. Ethylene gas, or artificial chemicals that mimic germination stimulants, can be applied to break dormancy and induce germination at times when no appropriate hosts are available. Alternatively, fumigants are applied to fields that kill seeds as well as other soil organisms. Raising the temperature of the soil by covering it with black plastic can also kill seeds.

Trap cropping, catch cropping, and crop rotation are other effective ways of managing the land. One option is to let the land lie fallow for 10 to 20 years to reduce the seed bank. This may not be an option where the demands on land use are heavy. Planting an unsuitable host is a more effective alternative that may allow the land to be used. Some alternative crops, called trap crops, induce seed germination but are not suitable hosts. A third option involves catch cropping. The land is seeded with susceptible hosts to induce germination of parasitic plants, but the crops are destroyed before the parasite can reproduce. Mixed cropping involves planting suitable and unsuitable hosts, and it has been demonstrated to reduce the impact of some witchweeds. The mechanisms, however, are not well understood.

Herbicides can also be used to control witchweeds. Herbicides applied directly to the soil after planting but before crop seedlings emerge can reduce or delay witchweed establishment. Systemic herbicides that are applied to the crop immediately after emergence may become translocated and concentrated in the parasites, killing them. The timing of application of systemic herbicides is critical because it must degrade before the crop begins to fruit. Some herbicides can be applied directly to the emerged witchweed to control them.

Damage to crops by witchweeds is minimized when nitrogen fertilizer is applied to host crops. This method has the added advantage of helping increase yields, but the benefits of using this method alone over a long period of time have not been demonstrated. The mechanisms that allow nitrogen fertilization to decrease the impact of the parasitic plants are not fully understood. It does seem clear, however, that on poor soils, the witchweed infestations increase in intensity as the quality of the soil decreases; the result is a feedback that continues until the land is almost useless and must be abandoned.

One of the most promising control options is the development of resistant varieties of crop plants through plant breeding. As new resistant varieties are found other methods of control become unnecessary. Resistant varieties of many crops have not been discovered, however, therefore this is a very active area of research.

Parasitic plants may pose the largest problem to rural, subsistence farmers. As pressures on land use increase, the ability to use crop rotation methods decreases. Furthermore, subsistence farmers often do not have the means to buy fertilizers, herbicides, or expensive resistant seeds. For these farmers, the feedbacks between poor soils and parasitic plant infestation may eventually cause the farmers to abandon plots of land and move elsewhere.

B. Mistletoes

Several species of mistletoes are economically important pests of fruit trees, ornamental trees, or timber. Most mistletoes have a similar ecology, so the impact and control measures are easy to generalize. The exceptions are dwarf mistletoes (genus *Arceuthobium*, Viscaceae) that have a much more highly developed endophytic component and inconspicuous, scaly leaves. Because of the highly developed endophyte, physical removal of dwarf mistletoes is more difficult than other mistletoes.

Most species of dwarf mistletoes are extremely host specific. For example, *A. douglasii* is a parasite of Douglas fir, and *A. tsugense* is a parasite of the western hemlock. The dwarf mistletoes are parasites of pine trees in North America, the Himalayas of India, Pakistan, and Bhutan, and in southwestern China. Dwarf mistletoes are common throughout the Rocky Mountains where large regions of forests are infested and the timber volume and quality are affected.

Two other genera in Viscaceae are notable for their economic impact. *Phoradendron* is a genus of about 190 species distributed throughout North and South America. Of these, the most important are *P. serotinum*, found on an extremely wide variety of hosts (but never conifers), and *P. piperoides*, found on cocoa in Costa Rica. *Viscum* is a genus of about 60 species that include *V. album* that has an extremely wide host range including fruit trees, pines, and poplars in Europe and persimmon in China. Another agricultural pest is *V. cruciatum* that parasitizes olives in the Middle East.

Many species of Loranthaceae cause economically significant loss. *Dendrophthoe falcata* is a pest of fruit trees in India and teak in Kerala. *Dendropthoe pentandra* parasitizes rubber and kapok in Indonesia. *Helixanthera mannii* is a pest of citrus and coffee in West Africa, and *H. parasitica* is a pest of citrus in the Himalayas. *Tapinanthus bangwensis* is an important pest of cocoa and cola in Ghana, while *T. globiferus* grows on coffee, citrus, and other fruit trees in Ethiopia. Some other genera that have economically important species are *Amyema*, *Englerina*, *Loranthus*, *Macrosolen*, *Oryctanthus*, *Phragmanthrea*, *Scurrula*, and *Struthanthus*.

Some resistant varieties of poplars, pines, and oaks do exist, but very little variability for resistance to mistletoe infections exists and, in general, breeding for resistance has not been a viable option. Biological control by insects, pathogens, or other mistletoes remains a possibility, but none of these methods have been successful. Instead, control of the leafy mistletoes involves pruning off infected limbs, or chemical control. Pruning infected limbs removes the mistletoe, but the limb must be cut below the point of the infection to remove the entire endophyte. This practice is extremely labor intensive and impractical on a large scale. With *Arceuthobium*, the endophyte is often extensive, and the limb must be severed 20 to 30 cm below the lowest parasite shoots. In extreme cases, it may be more cost-effective to burn the forest. In less extreme cases, early harvest followed by controlled burns may reduce the mistletoes. Some successful chemical control methods have been found. The leaves of mistletoes may be sprayed with ethephon once the host drops its leaves. Alternatively, herbicides can be injected into infected limbs below the point of infection, accumulate in mistletoes, and eventually kill them without severely affecting the host.

C. Dodders

The economic impact of *Cuscuta* species is greatest in forage crops such as alfalfa and clover; however, crops like citrus, coffee, peach, litchi, flax, linseed, and other crops as well as ornamental plants all suffer from dodder attack. The most economically important dodder is *Cuscuta campestris*. This species attacks alfalfa and has been shown to reduce forage yield by as much as 57% over a 2-year period. It also affects Niger seed in India and many vegetable and flower crops. Of the 20 species of *Cassytha*, the most damaging is *C. filiformis*. It is distributed throughout the tropics and parasitizes a wide range of hosts. It is a problem on citrus in India and Tanzania and on *Pinus massonia* in China.

Mechanical control methods for *Cuscuta* include hand pulling, crop rotation, burning, delaying planting until after *Cuscuta* have germinated, or deep ploughing to reduce the seed bank. Few resistant varieties of crops are known. Chemical control methods include fumigants to eliminate the seed bank, herbicides applied to the soil to prevent seedling growth, and herbicide application after seeds have germinated to prevent establishment. Some methods of biological control have been established with insects and pathogens, but the scope of these has been limited.

XI. CONCLUSIONS

Parasitic plants have been understudied, but recent studies have improved our understanding of their evolution, ecology, molecular biology, and physiology. Haustorial parasitism has evolved independently at least 12 times within angiosperms and, surprisingly, complete loss of photosynthesis has occurred multiple times within some lineages. Most parasitic plants exhibit near-normal levels of photosynthesis like their autotrophic ancestors, whereas others are incapable of photosynthesis such as the derived endophytic holoparasites. Study of parasites that photosynthesize at intermediate levels will be important to understand the evolution of holoparasitism. Development of sophisticated molecular methods will be important for understanding the bases of haustorial development and should ultimately enhance efforts to control economically important crop pests. Further ecological and physiological studies aimed at understanding host-parasite interactions will also enhance efforts to control parasitic weeds.

Acknowledgments

We thank the NSF (grants DEB-91029258, DBI-9604814, and DEB-9811362 to C. W. D.) for funding our research on parasitic plants, and Joel McNeal for critical reading of the manuscript. Other important contributions of data, samples, or discussion, were also made by other members and friends of the dePamphilis lab, past and present, and by numerous botanical gardens and botanists with an interest in parasitic plants.

See Also the Following Articles

COEVOLUTION • PARASITISM • PARASITOIDS

Bibliography

Atsatt, P. R. (1970). Hemiparasitic Flowering plants: Phenotypic canalization by hosts. *Nature* **225**, 1161–1163.

dePamphilis. C. W. (1995). Genes and genomes. In *Parasitic Plants*. (M. C. Press and J. D. Graves., Eds.), pp. 177–205. Chapman and Hall, London.

Finlay, R. D., and Read, D. J. (1986). The structure and function of the vegetative mycelium of ectomycorrhizal plants. I: Translocation of C-labelled carbon between plants interconnected by a common mycelium. *New Phytologist* **103**, 143–156.

Graham, B. F. Jr., and Bormann, F. H. (1966). Natural root grafts. *Bot. Rev.* **32**, 255–292.

Hawksworth, F. G. andWeins, D. (1972). Biology and classification of dwarf mistletoes (*Arceuthobium*). Agriculture handbook no. 401, U.S. Government Printing Office, Washington D.C.

Kuijt, J. (1969). *The Biology of Parasitic Flowering Plants*. University of California Press, Berkeley, CA.

Marvier, M. A. (1996). Parasitic plant-host interactions: Plant performance and indirect effects on parasite-feeding herbivores. *Ecology* 77, 1398–1409.

Musselman, L. J. (1980). The biology of *Striga, Orobanche,* and other root-hemiparasitic weeds. *Annual Review of Phytopathology* 18, 463–489.

Musselman, L. J. (1987). *Parasitic weeds in agriculture Vol. I: Striga.* CRC Press, Boca, Raton, FL.

Nickrent, D. L. (2000, January) The parasitic plant connection, http://www.science.siu.edu/parasitic-plants/index.html.

Nickrent. D. L., Duff, R. J., Colwell, A. E., Wolfe, A. D., Young, N. D., Steiner, K. E., and dePamphilis, C. W. (1998). Molecular phylogenetic and evolutionary studies of parasitic plants. In *Molecular Systematics of Plants, II,* (D. Soltis, P. Soltis, and J. Doyle, Eds.), Chapter 14.

Parker, C., and Riches, C. R. (1993). *Parasitic Weeds of the World: Biology and Control.* CAB International, Wallingford, Oxon OX10 8DE, UK.

Press, M. C., and Graves, J. D. Graves (Eds.) (1995). *Parasitic Plants.* Chapman & Hall, London, UK.

Stewart, G. R., and Press, M. C. 1990. The physiology and biochemistry of parasitic angiosperms. *Annual Review of Plant Physiology and Plant Molecular Biology* 41, 127–151.

Yoder, J. I. (1997). A species-specific recognition system directs haustorium development in the parasitic plant *Triphysaria* (*Scrophulariaceae*). *Planta* 202, 407–413.

Young, N. D., Steiner, K., and dePamphilis, C. W. (1999). The evolution of parasitism in Scrophulariaceae/Orobanchaceae: Plastid gene sequences refute an evolutionary transition series. *Ann. Missouri Bot. Gard.* 86, 876–893.

HERBACEOUS VEGETATION, SPECIES RICHNESS IN

J. P. Grime
The University of Sheffield

GLOSSARY

dominant species Species that have the greatest influence on ecosystem structure and function by virtue of their abundance, biomass, or coverage.

mass ratio hypothesis Hypothesis stating that ecosystem processes are largely determined by the dominant contributors of the overall plant biomass, that is, dominant species will exert greater influence on processes than will subordinate species.

species richness Number of species present in a given habitat or ecosystem.

subordinate species Species that have a minor influence on ecosystem structure and function, presumably because of their lesser abundance and biomass compared to dominant species.

transient species Species that are present as scattered seedlings or small immature individuals; many of these species occur as dominant or subordinate species in neighboring vegetation associated with different environmental conditions or management regimes.

IN THE SECOND HALF of the twentieth century, reductions in plant diversity have been observed in many types of ecosystems. One of the best-known and most lamented of these losses has been associated with the widespread disappearance and degradation of ancient species-rich meadows and pastures in western Europe. However, from census data collected in several countries (Ratcliffe, 1984; Thompson, 1994; Thompson and Jones, 1999), we have learned not to regard this phenomenon in isolation from processes taking place in landscapes as a whole; the fate of such highly prized communities is merely one conspicuous element in a widespread attrition affecting both species-rich and species-poor ecosystems.

I. INTRODUCTION

Ecological research into losses in plant diversity has a long history, but since 1990 the agenda has shifted from a primary concern with its mechanisms to an assessment of the extent to which losses impair the functioning of ecosystems and reduce their utility to humankind. It is relatively easy to manipulate the spe-

Encyclopedia of Biodiversity, Volume 3

cies composition and diversity of herbaceous communities in both natural and model ecosystems, and experimenters have applied themselves with vigor to this new field of enquiry. The objectives in this research are twofold. First, we need to identify where and how losses in diversity affect ecosystem functions such as primary productivity, carbon storage, mineral nutrient cycling, and ecosystem resistance and resilience. Second, we require information on the plant constituents necessary for successful ecosystem restorations; we need to establish which plant species are irreplaceable and what population sizes and pools of genetic variation are required for their persistence.

As a participant in both the current and the preceding phases of research on plant diversity, I have become convinced of the value of connecting theory and data from the earlier phase of research into the causes of plant diversity to the current (post-1990) effort to examine its consequences for ecosystems. The main objective of this article is to identify these connections. An additional purpose is to place in perspective a current debate in which divergent conclusions have been drawn concerning the importance of high plant species richness as an immediate controller of ecosystem properties. This debate provides a useful introduction to the issues that will be considered later.

II. PLANT TRAITS OR SPECIES RICHNESS?: THE CURRENT DEBATE

A. Effects of Plant Traits on Ecosystem Properties

Although the majority of investigations in plant ecophysiology have sought to explain how vegetation is determined by environment, there has also been widespread recognition of feedbacks in which plants affect their environments and influence ecosystem properties. This idea is implicit in the writings of the founders of ecosystem theory (Odum, 1963, 1969) and more recently has found expression in the search for plant functional types (Grime, 1979; Chapin, 1980; Smith *et al.*, 1996), in which specific plant community and ecosystem properties have been attributed directly to particular plant traits such as potential growth rate, palatability to herbivores, resistance to fire, litter quality, and seed persistence. This approach has been extended further by studies in which the predictive value of plant traits has been tested by examining ecosystem responses to perturbation. An early example is the investigation by Leps *et al.* (1982) in which differences between two neighboring grasslands in their resistance and resilience following the severe drought of 1975 in Czechoslovakia were predicted from the life histories and potential growth rates of the main component plant species. More comprehensive tests were conducted by MacGillivray *et al.* (1995), in which responses to frost, drought, and fire in five contrasted grasslands were successfully predicted from a set of plant traits measured in a laboratory screening program (Grime *et al.*, 1997).

B. The Mass Ratio Hypothesis

It is important to note that in tests such as those conducted by MacGillivray *et al.* (1995), predictions are weighted according to the abundance of each plant species in the vegetation; this is founded on the assumption that the extent to which the traits of a species affect ecosystem properties is likely to be strongly related to the contributions of the species to processes such as photosynthesis, mineral nutrient capture, transpiration, and provision of substrates exploited by herbivores and decomposers. There is a clear implication here that ecosystem processes are determined to a very large extent by the characteristics of the dominant contributors to the plant biomass. As a corollary of this "mass ratio hypothesis" (Grime, 1998), we would not expect minor contributors to the vegetation to exert strong effects on ecosystem properties.

Is it reasonable to conclude that ecosystem functions can be sustained by inputs originating from only a few dominant plant species? Some of the most extensive ecosystems contain very few plant species; this is particularly obvious in grasslands and heathlands on acidic soils, and here the conclusion is inescapable that controls on ecosystem processes by vascular plants are mediated through very few species. In herbaceous vegetation on soils of higher base status, analysis is complicated by the presence of more species, and in old calcareous grasslands the densities of herbs can rise to 30–40 species per m^2. Even here, however, it is interesting to note that when quantitative studies are made by harvest methods (Al-Mufti *et al.*, 1977) or point analysis (Mitchley and Grubb, 1986; see also Fig. 1), it is evident that in comparison with the canopy dominants many of the subordinate species of such communities account for a very small proportion of the biomass. There must be considerable doubt as to whether such minor components can, even collectively, exercise *immediate* effects on properties such as productivity, carbon storage, and water relations.

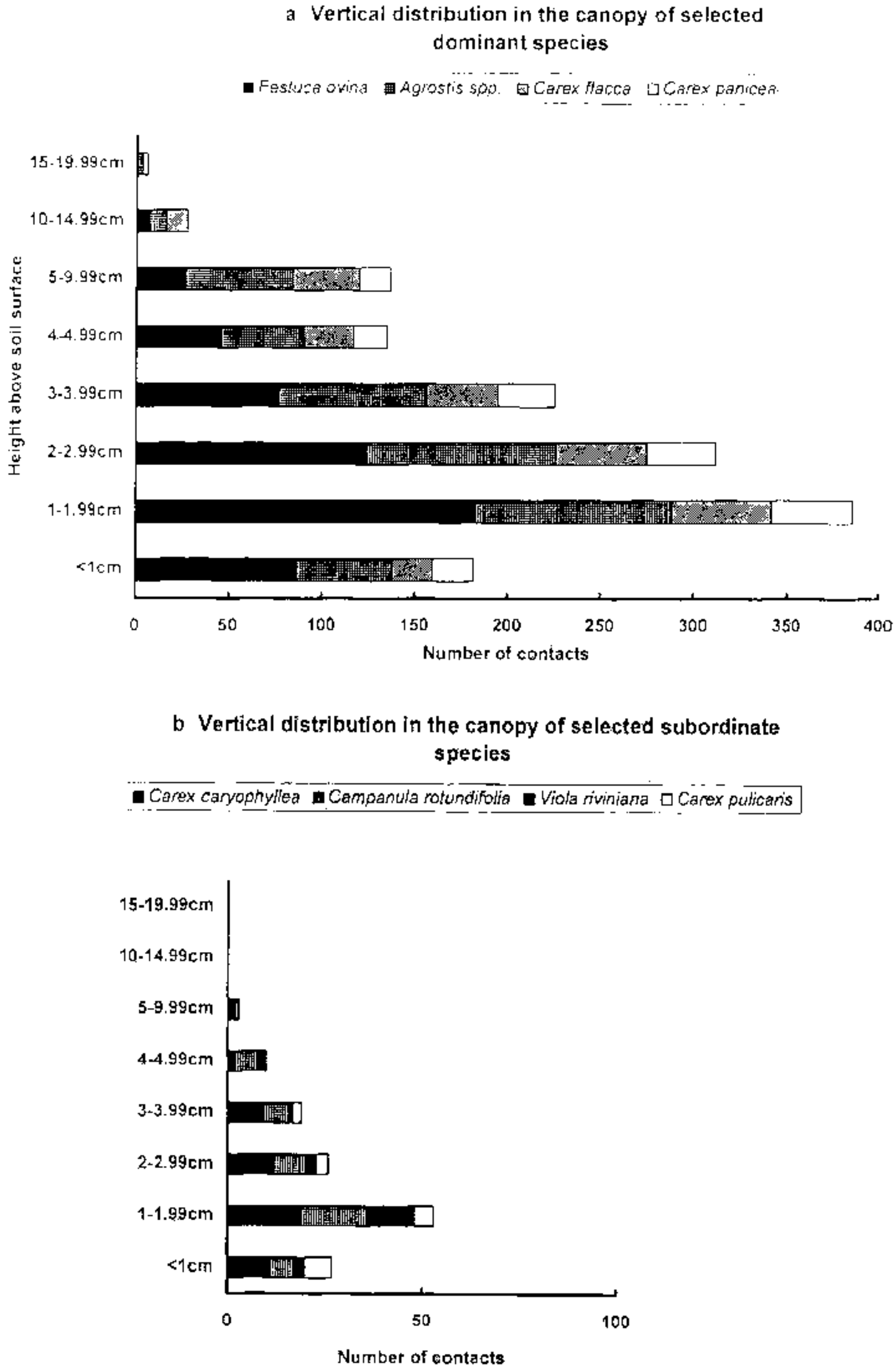


FIGURE 1 Vertical distribution in the leaf canopy of four dominant and four subordinate component species in an ancient limestone pasture at Buxton, North Derbyshire, England. Canopy distribution in June was estimated by measuring contacts with 375 randomly distributed, vertical pins. (S. H. Hillier, unpublished data.)

C. Effects Associated with High Species Richness

Against a theoretical background in which the traits of dominant plants were widely suspected to be acting as the overriding controllers of ecosystem properties, considerable interest and controversy were generated when, in 1994, two papers appeared purporting to demonstrate the immediate benefits to ecosystem properties arising from high species richness in experimental plant assemblages (Naeem *et al.*, 1994; Tilman and Downing, 1994). In each case it was suggested that benefits arose in the species-rich mixtures from the presence of a wider range of morphologies and physiologies, generating either complementary and more complete exploitation of resources (Naeem *et al.*, 1994) or conferring resistance and resilience in the face of an extreme event (Tilman and Downing, 1994). Interest in these publications extending beyond the realm of ecology was stimulated by commentaries (Kareiva, 1994, 1996) suggesting that studies of this kind provided a justification for the conservation of species-rich ecosystems.

Subsequently, doubts have been cast on the validity of the conclusions drawn by the authors of the two papers and these have been reviewed in detail elsewhere (Givnish, 1994; Aarssen, 1997; Garnier *et al.*, 1997; Huston, 1997; Grime, 1998; Hodgson *et al.*, 1998; Holmes, 1998). It appears that, in both cases, ecosystem properties attributed to high species richness were in reality due to the presence in the more diverse communities of dominant species with traits attuned, respectively, to high productivity (Naeem *et al.*, 1994) and drought (Tilman and Downing, 1994). More recent experiments (Wardle *et al.*, 1997; Hooper and Vitousek, 1997; Tilman *et al.*, 1997) have failed to provide convincing support for the effects of high species richness on ecosystem functions; the most parsimonious explanation for the data presented in all of these papers is that the ecosystem properties examined were controlled by the functional traits of a relatively small number of species accounting for a high proportion of the total plant biomass.

The alacrity with which some commentators and conservation enthusiasts seized upon supposed evidence of the direct benefits of species richness to ecosystems emphasizes the need for caution in a complex research field. It also brings the risk that exposure of false conclusions may lead to a sceptical response if and when more substantial evidence of the beneficial effects of species richness is presented. Perhaps the greatest danger arising from the current state of research on this subject is the almost exclusive emphasis on *immediate* benefits of richness on ecosystem functions. Elsewhere (Grime, 1998) it has been proposed that the most potent effects of declining biodiversity are likely to become evident only in the longer term. To explore this argument it is necessary to refer to much older research on plant diversity that has been largely ignored in the recent burst of new experimental work.

III. COMPONENTS OF SPECIES RICHNESS: DOMINANTS, SUBORDINATES, AND TRANSIENTS

A. Dominants and Subordinates

Early efforts to describe and interpret herbaceous vegetation (e.g., Clements, 1905; Tansley, 1939; Ramenskii, 1938) involved listing of the plants present in selected stands of vegetation and estimating the abundance of each species. Even before the widespread adoption of experimental approaches there was a keen awareness of the potential of certain species to occupy a high proportion of the plant biomass and to control the abundance and fitness of other minor contributors. A classic example is the rapid expansion of tall coarse grasses and the coincident suppression of small herbs noted by Tansley and Adamson (1925) in their study of the consequences of excluding rabbits from chalk grassland.

A more quantitative approach followed as plant ecologists adopted stricter sampling methods and measured the relative abundance of plant species by harvesting, sorting, and weighing vegetation samples. This allowed more formalized attempts to examine the functional relationships between dominant and subordinate members of plant communities (Whittaker, 1965; McNaughton, 1978). Subsequently, it was recognized that the potential to dominate vegetation could be associated with specific plant traits such as canopy height, lateral spread, and the capacity to project shoots forcefully through litter and herbaceous cover (Al-Mufti *et al.*, 1977; Sydes and Grime, 1981a,b; Grubb *et al.*, 1982; Campbell *et al.*, 1992). A further step in defining the functional differences between potentially dominant and subordinate members of plant communities became possible through the application of foraging theory to plants. From a study of the development of the roots and shoots of isolated plants exposed to standardized patchy environments (Campbell *et al.*, 1991), it became possible to predict the dominance hierarchy that developed when eight herbaceous species were grown together in an experimental assemblage (Fig. 2). In this investigation it was found that dominance was associated with relatively imprecise foraging both above and below ground. Such coarse-grained foraging allowed the exploitation of a large volume of habitat but was associated with an imprecise concentration in resource-rich sectors. In marked contrast, subordinate species were characterized by precise but local placements of roots and shoots in resource-rich patches, a foraging pattern permitting temporary coexistence with dominants but apparently committing them to a minor status in the community and leaving them vulnerable to extinction in circumstances where the dominants continued to enjoy unrestricted growth.

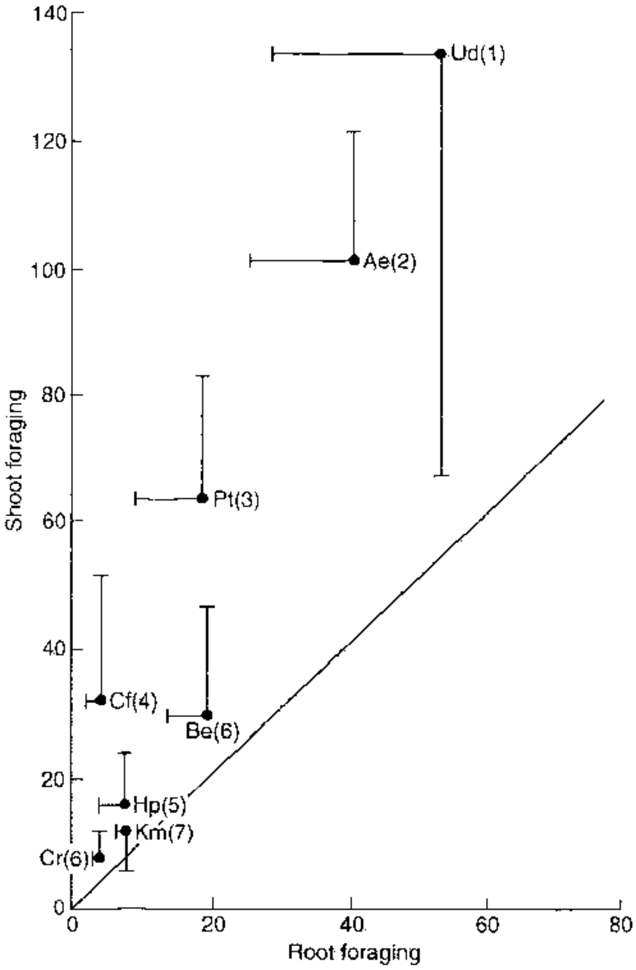


FIGURE 2 An examination of the relationship between root and shoot responses to resource heterogeneity in eight herbaceous species of contrasted ecology. A description of the methods used to expose the plants to resource patchiness is provided in Campbell *et al.* (1991). Scales of foraging by the roots and shoots in the foraging assays are expressed at the respective increments of biomass (mg) to two undepleted quadrants, which in both assays constitute 50% of the available volume. The vertical and horizontal bars indicate 95% confidence limits. The numbers in parentheses refer to the species ranking in a conventional competition experiment in which all eight species were grown together in an equiproportional mixture on fertile soil for 16 weeks. Note the consistent tendency for the scale of leaf canopy adjustment to exceed that of the root system. This of course arises from the freedom of movement of leaves in air and the encasement of roots in soil. Key to species: Ae, *Arrhenatherum elatius*, Be, *Bromus erectus*; Cf, *Cerastium fontanum*; Cr, *Campanula rotundifolia*; Hp, *Hypericum perforatum*; Km, *Koeleria macrantha*; Pt, *Poa trivialis*; Ud, *Urtica dioica*.

These findings raise important questions concerning the ecology and evolution of the very large numbers of plant species that consistently occupy subordinate positions within the hierarchies of herbaceous communities. Are these plants merely "also-rans" in the struggle for existence or are there fitness benefits associated with playing a minor part in plant communities? What is the precise role of subordinates in the long-term functioning and dynamics of ecosystems? These are not trivial questions in view of the fact that it is the subordinates rather than dominants that account numerically for the existence of high species richness in plant communities. Detailed attempts have been made to address these issues (Grime, 1987, 1998); here argument will be reduced to the following propositions:

1. Many subordinates achieve widespread occurrence because they are capable of exploiting similar niches associated with different dominant species. For example, *Poa trivialis*, the commonest vascular plant species of Lowland Britain, occupies shaded microhabitats beneath tall grasses, broad-leaved herbs, and a variety of shrubs and trees.
2. In many species-rich communities the risks to subordinates of competitive exclusion remain low because potential dominants are restricted in vigor by environmental and biotic factors (e.g., mineral nutrient stress, drought, grazing, fire).
3. Because they do not experience the costs in support materials and construction time incurred during the building of robust root and shoot architectures, subordinates enjoy a temporary advantage when communities are subject to episodes of biomass removal. Where damage is catastrophic, as in tree felling and coppicing or the ploughing and burning of grasslands (Skutch, 1929; Marks, 1974; Platt, 1975; Bormann and Likens, 1979; Pickett and White, 1985; Pons, 1989), this can lead to a temporary but massive expansion of subordinates until such time as the dominants recover.

We may conclude, therefore, that many subordinate members of plant communities achieve high levels of fitness, particularly where natural or human-inspired interventions restrict the vigor of potential dominants either continuously or intermittently. Later in this article, it will be suggested that, in the long term, this rather opportunistic ecology displayed by subordinates may be sufficient to allow them an indirect but crucial involvement in the determination of ecosystem properties.

B. Transients

Many published records of the species composition of vegetation are incomplete. This can arise from cursory sampling and recording or from a deliberate policy of discounting minor constituents. Where the objective is to recognize recurring plant communities, data analysis often involves procedures that exclude species of low frequency or inconsistent occurrence in vegetation samples. The result of these various decisions can be a divergence between published data and field reality, and it has been suggested (Grime, 1998) that important information relevant to the long-term dynamics of ecosystems may be lost. To examine this hypothesis it is necessary to refer to surveys in which a strenuous attempt has been made to include all the plant species present in the vegetation.

During the period 1965–1975, a series of vegetation surveys was conducted to produce an inventory of the herbaceous plant communities of the Sheffield region in north-central England (see Grime *et al.*, 1988, for details). In each of approximately 10,000-m^2 quadrats, a complete list of species was made and the frequency of each species was recorded. When the resulting data were examined it became apparent that, in addition to dominants and subordinates, the majority of the samples contained species that were represented only as scattered seedlings and small immature individuals. Most of these species were present as dominants or subordinates in neighboring vegetation associated with different environmental conditions or management regimes. Examples of this phenomenon are provided in Table I for three contrasted sets of vegetation samples; in each case it is evident that the plant communities examined were harboring juveniles of species that occur as persistent, reproductive populations elsewhere in the landscape in vegetation of a different type.

Therefore, it seems reasonable to conclude that in addition to the dominant and subordinate members of a community, a third contributor to species richness can be identified in the form of transients originating either from the seed rain from the surrounding landscape or from seed banks occurring as a legacy of previous vegetation types that occupied the site.

IV. DECLINING SPECIES RICHNESS AND ECOSYSTEM REASSEMBLY

If immediate control of the functional properties of an ecosystem rests with dominant plants and if species-richness depends on the numbers of subordinates and transients, it is pertinent to ask "Does a decline in species richness matter?" To address this question, it is

TABLE I

Ecological Classification of the Species Recorded in Three Distinctive and Highly Contrasted Habitats Sampled Widely in an Area of 2400 km^2 around Sheffield, United Kingdom[a]

Sampled habitat	Number of m^2 samples	Woodland species	Grassland species	Arable species	Others[b]	Total
Woodland on limestone	51	65	23	1	5	94
Meadows	40	7	64	8	0	79
Cereal arable	55	5	38	69	2	114

[a] All the species encountered within a particular habitat were classified in terms of their primary habitat (columns 3–6). Details of the sampling, recording, and habitat classification procedures are provided in Grime *et al.* (1988).

[b] Includes species primarily associated with wetland or skeletal habitats (cliffs, walls, and rock outcrops).

necessary to consider the significance of losses of subordinates and transients in the long-term dynamics of plant communities and ecosystems.

A. Filter Effects of Subordinates?

One mechanism whereby losses in subordinates could affect ecosystems is through alteration of the filter controlling the recruitment, identity, and relative abundance of dominants. To review the opportunities for subordinates to control the admission of dominants into communities it is necessary to consider the long-term dynamics of vegetation and the regenerative phases in the life cycles of dominants. Studies of vegetation succession conducted earlier this century (e.g., Watt, 1925, 1947) established that continued dominance by particular species is frequently determined by the success of seedling or vegetative reestablishment following disturbance events. As already described here, often the early course of events following a disturbance is a temporary expansion in the cover and vigor of subordinates. This is most obvious in forest clearings where a dense, low cover of shrubs, herbs, and bryophytes characterizes the environment of regenerating trees (Watt, 1925; Skutch, 1929; Marks, 1974; Bormann and Likens, 1979), but similar phenomena have been described for grasslands and heathlands (Oosting, 1942; Keever, 1950; Hillier, 1990). Establishment following disturbances involves complex interactions of seedlings and vegetative shoots with substratum conditions, and contributions to the ground cover by subordinate plants may be expected to have both positive and negative effects (Pickett and White, 1985). Benefits to establishment have been described in circumstances where seedlings survive in the shelter afforded by low-growing shrubs, herbs, and bryophytes. Detrimental effects of shrub, herbaceous, and bryophyte cover on the establishment of grassland and forest dominants have been observed (Niering and Goodwin, 1962; Webb *et al.*, 1972; Pons, 1989), and it is widely recognized that many small-seeded herbs, trees, and shrubs are incapable of establishment in a closed cover of vegetation.

We may deduce, therefore, that there is a potential for subordinate members of a plant community to act as a filter in selecting between different potential dominants during the early phases of recolonization following a disturbance event. Selection could operate on the basis of variation in the seed reserves of dominants and on the capacity of their seedlings to penetrate a low canopy. The filter might also discriminate between dominants that rely on rapid emergence and those that regenerate by persistent juveniles. Subordinates could also control regenerating dominants through more indirect mechanisms, such as provision of sites in which seed predation is reduced, or through more complex phenomena such as the maintenance of critical pests, pathogens, herbivores, or mutualists.

The filter roles that subordinates may play during ecosystem reassembly require confirmation by carefully designed, long-term experiments.

B. Founder Effects of Transients?

If, as suggested earlier in this article, the sources of the transients are seed banks in the soil and the seed rain from the surrounding landscape, it would appear that they are an index of the pool of potential colonizing species at each site. On this basis, a diversity of transients signifies a high probability that, in the event of habitat disturbance or changes in management, there will be a rapid ingress of different plant functional types, some of which may be capable of exploiting the new conditions. An obvious example is the benefit to woodland development where an abandoned grassland already contains a diverse assortment of tree seedlings.

Current losses in biodiversity in Europe and in many other parts of the world are taking place in a complex landscape mosaic that is continuously disturbed by natural events and by urbanization, arable cultivation, forestry, and various forms of grassland management. Ecosystem sustainability depends in part on the continuous movement of populations and the reassembly of vegetation types and ecosystems. The extent to which communities and ecosystems are rapidly reconstituted is likely to be related to the reservoir of colonizers, many of which should be detectable prior to disturbance as transient constituents of the existing vegetation. As Egler (1954) recognized, we may suspect that the speed and completeness with which ecosystem reassembly occurs will depend on early colonization by appropriate dominants and subordinates; late arrivals will be delayed in their establishment and some may be excluded completely (Keever, 1950; Niering and Goodwin, 1952; Platt, 1975). It is not difficult to envisage how circumstances could then arise whereby efficiently dispersed plant species with "poor fit" to habitat and management conditions could assume dominance with damaging consequences for ecosystem function. There is an urgent need to discover the extent to which failure in the processes of plant dispersal and ecosystem reassembly can be predicted from the decline through time in the density and species richness of transients in plant communities.

V. CONCLUSIONS

Currently the balance of evidence (Huston, 1997; Grime, 1998) is shifting toward the mass ratio hypothesis and against the proposition that species richness itself controls the immediate functioning of ecosystems. However, this does not mean that losses of plant diversity should be ignored. The priority in the next phase of research on declining plant diversity should be to consider its long-term consequences for ecosystem structure and function. Losses in species richness may be associated with less obvious impacts that operate through failures in filter and founder effects. A progressive loss of ecosystem functions may be predicted in circumstances where vegetation patch dynamics and ecosystem reassembly continue against the background of a declining pool of colonizing propagules. The effects on the recruitment of dominants, rather than the immediate consequences of declining richness per se, deserve our curiosity and attention.

See Also the Following Articles

ECOSYSTEM FUNCTION, PRINCIPLES OF • HERBICIDES • PLANT BIODIVERSITY, OVERVIEW

Bibliography

Aarssen, L. W. (1997). High productivity in grassland ecosystems: Effected by species diversity of productive species? *Oikos* 80, 183–184.

Al-Mufti, M. M., Sydes, C. L., Furness, S. B., Grime, J. P., and Band, S. R. (1977). A quantitative analysis of shoot phenology and dominance in herbaceous vegetation. *J. Ecol.* 65, 759–791.

Bormann, F. H., and Likens, G. E. (1979). *Pattern and Process in a Forested Ecosystem.* Springer-Verlag, Berlin.

Campbell, B. D., Grime, J. P., and Mackey, J. M. L. (1991). A tradeoff between scale and precision in resource foraging. *Oecologia* 87, 532–538.

Campbell, B. D., Grime, J. P., and Mackey, J. M. L. (1992). Shoot thrust and its role in plant competition. *J. Ecol.* 80, 633–641.

Chapin, F. S. (1980). The mineral nutrition of wild plants. *Annu. Rev. Ecology and Systematics* 11, 233–260.

Clements, F. E. (1905). *Research Methods in Ecology.* University Publishing Company, Wisconsin.

Garnier, E., Navas, M., Austin, M. P., Lilley, J. M., and Gifford, R. M. (1997). A problem for biodiversity–productivity studies: How to compare the productivity of multispecific plant mixtures to that of monocultures? *Acta Oecologica* 18, 657–670.

Givinish, T. J. (1994). Does diversity beget stability? *Nature* 371, 113–114.

Grime, J. P. (1979). *Plant Strategies and Vegetation Processes.* John Wiley & Sons, Chichester, United Kingdom.

Grime, J. P. (1987). Dominant and subordinate components of plant communities—Implications for succession, stability and diversity. In: *Colonisation, Succession and Stability* (A. Gray, P. Edwards, and M. Crawley, eds.), pp. 413–428. Blackwell Scientific Publications, Oxford, United Kingdom.

Grime, J. P. (1998). Benefits of plant diversity to ecosystems: Immediate, filter and founder effects. *Journal of Ecology*, 902–910.

Grime, J. P., Hodgson, J. G., and Hart, R. (1988). Comparative Plant Ecology: a Functional Approach to Common British Species. Unwin Hyman, London.

Grime, J. P., *et al.* (1997). Integrated screening validates primary axes of specialisation in plants. *Oikos* 79, 259–281.

Grubb, P. J., Kelly, D., and Mitchley, J. (1982). The control of relative abundance in communities of herbaceous plants. In *The Plant Community as a Working Mechanism*, (E. Newman, ed.), British Ecological Society No. 1, pp. 77–97. Blackwells, Oxford, United Kingdom.

Hillier, S. H. (1990). Gaps, seed banks and plant species diversity in calcareous grasslands. *Calcareous Grasslands: Ecology and Management* (S. H. Hillier, D. W. H. Walton and D. A. Wells, eds.), pp. 57–66. Bluntisham Books, Huntingdon.

Hodgson, J. G., Thompson, K., Bogaard, A., and Wilson, P. (1998). Does biodiversity determine ecosystem function? The Ecotron experiment reconsidered. *Functional Ecology*, 843–848.

Holmes, B. (1998). Life support: Why bother to save every last species on the planet? *New Scientist* 2147, 30–34.

Hooper, D., and Vitousek, P. M. (1997). The effects of plant composition and diversity on ecosystem processes. *Science* 277, 1302–1305.

Huston, M. A. (1997). Hidden treatments in ecological experiments: Evaluating the ecosystem function of biodiversity. *Oecologia* 110, 449–460.

Kareiva, P. (1994). Diversity begets productivity. *Nature* 368, 686–687.

Kareiva, P. (1996). Diversity and sustainability on the prairie. *Nature* 379, 673–674.

Keever, C. (1950). Causes of succession on old fields of the Piedmont, North Carolina. *Ecological Monographs* 20, 229–250.

Leps, J., Osbornova-Kosinova, J., and Rejmanek, M. (1982). Community stability, complexity and species life-history strategies. *Vegetatio* 50, 53–63.

MacGillivray, C. W., Grime, J. P., and the ISP team. (1995). Testing predictions of resistance and resilience of vegetation subjected to extreme events. *Functional Ecol.* 9, 640–649.

Marks, P. L. (1974). The role of pin cherry (*Prunus pensylvanica* L.) in the maintenance of stability in northern hardwood ecosystems. *Ecol. Monographs* 44, 73–88.

McNaughton, S. J. (1978). Stability and diversity of ecological communities. *Nature* 274, 251–253.

Mitchley, J., and Grubb, P. J. (1986). Control of relative abundance of perennials in chalk grassland in southern England. I. Constancy of rank order and results of pot- and field-experiments on the role of interference. *J. Ecol.* 74, 1139–1166.

Naeem, S., Thompson, L. J., Lawler, S. P., Lawton, J. H., and Woodfin, R. M. (1994). Declining biodiversity can alter the performance of ecosystems. *Nature* 368, 734–737.

Niering, W. A., and Goodwin, R. H. (1962). Ecological studies in the Connecticut Arboretum Natural Area. I. Introduction and survey of vegetation types. *Ecology* 43, 41–54.

Odum, E. P. (1963). *Ecology.* Holt, Rinehart and Winston, New York.

Odum, E. P. (1969). The strategy of ecosystem development. *Science* 164, 262–270.

Oosting, H. J. (1942). An ecological analysis of the plant communities of Piedmont, North Carolina. *American Midland Naturalist* 28, 1–126.

Pickett, S. T. A., and White, P. S. (eds.). (1985). *The Ecology of Natural Disturbance and Patch Dynamics.* Academic Press, Orlando, Florida.

Platt, W. (1975). The colonisation and formation of equilibrium plant species associations on badger disturbances in a tall-grass prairie. *Ecol. Monographs* 45, 285–305.

Pons, T. L. (1989). Dormancy, germination and mortality of seeds in heathland and inland sand dunes. *Acta Botanica Neerlandica* 38, 327–335.

Ramenskii, L. G. (1938). *Introduction to the Geobotanical Study of Complex Vegetations.* Selkzgiz, Moscow.

Ratcliffe, D. A. (1984). Post-medieval and recent changes in British vegetation: The culmination of human influence. *New Phytologist* 98, 73–100.

Skutch, A. F. (1929). Early stages of plant succession following forest fire. *Ecology* 10, 177–190.

Smith, T. M., Shugart, H. H., and Woodward, F. I. (1996). *Plant Functional Types: Their Relevance to Ecosystem Properties and Global Change.* Cambridge University Press, Cambridge, United Kingdom.

Sydes, C., and Grime, J. P. (1981a). Effects of tree leaf litter on herbaceous vegetation in deciduous woodland. I. Field investigations. *J. Ecol.* 69, 237–248.

Sydes, C., and Grime, J. P. (1981b). Effects of tree leaf litter on herbaceous vegetation in deciduous woodland. II. An experimental investigation. *J. Ecol.* 69, 249–262.

Tansley, A. G. (1939). *The British Islands and Their Vegetation.* Cambridge University Press, Cambridge, United Kingdom.

Tansley, A. G., and Adamson, R. S. (1925). Studies of the vegetation of the English Chalk. III. The chalk grasslands of the Hampshire–Sussex border. *J. Ecol.* 13, 177–223.

Thompson, K. (1994). Predicting the fate of temperate species in response to human disturbance and global change. In *Biodiversity, Temperate Ecosystems and Global Change* (T. J. B. Boyle and C. E. B. Boyle, eds.). Springer-Verlag, Berlin.

Thompson, K., and Jones, A. (1999). Human population density and prediction of local plant extinction in Britain. *Conservation Biol* 15, 1–6.

Tilman, D., and Downing, J. A. (1994). Biodiversity and stability in grasslands. *Nature* 367, 363–365.

Tilman, D., Knops, J., Wedin, D., Reich, P., Ritchie, M., and Siemann, E. (1997). The influence of functional diversity and composition on ecosystem processes. *Science* 277, 1300–1302.

Wardle, D. A., Zackrisson, O., Hornberg, G., and Gallet, C. (1997). The influence of island area on ecosystem properties. *Science* 277, 1296–1299.

Watt, A. S. (1925). On the ecology of the British beechwoods with special reference to their regeneration. Part II. Sections II and III. The development and structure of beech communities on the Sussex Downs. *Journal of Ecology* 13, 27–73.

Watt, A. S. (1947). Pattern and process in the plant community. *Journal of Ecology* 35, 1–22.

Webb, L. J., Tracey, J. G., and Williams, W. T. (1972). Regeneration and pattern in the subtropical rain forest. *Journal of Ecology* 60, 675–695.

Whittaker, R. H. (1965). Dominance and diversity in land plant communities. *Science* 147, 250–260.

HERBICIDES

Jodie S. Holt

University of California, Riverside

GLOSSARY

herbicide Chemical used to suppress or kill plants, or to severely interrupt their normal growth processes.

herbicide resistance Inherited ability of a plant to survive and reproduce following exposure to a dose of herbicide normally lethal to the wild type. In a plant, resistance may be naturally occurring or induced by techniques such as genetic engineering or selection of variants produced by tissue culture or mutagenesis.

herbicide tolerance Inherent ability of a species to survive and reproduce after herbicide treatment. This implies that there was no selection or genetic manipulation to make the plant tolerant; it is naturally tolerant.

integrated weed management Approach for suppressing weeds that combines information on the biology and ecology of the weed with all available control technologies so that no one method is used exclusively.

postemergence herbicide Herbicide applied after the emergence of the specified weed or crop.

preemergence herbicide Herbicide applied to the soil prior to the emergence of the specified weed or crop.

selectivity Phenomenon in which some plants are killed with doses of herbicides that have little or no effect on other plants.

weed Plant that interferes with the growth of desirable plants and is unusually persistent and pernicious. Weeds negatively affect human activities and as a result are undesirable.

weed control Reducing or suppressing weeds in a defined area to an economically acceptable level without necessarily eliminating them.

HERBICIDES ARE USED IN AGRICULTURAL, AQUATIC, FOREST, AND WILDLAND ECOSYSTEMS to reduce the density of unwanted vegetation (weeds) and to permit the growth of desirable species. Over time, use of herbicides in agroecosystems reduces weed density and may also select species that are adapted to the particular chemicals, thereby reducing weed species diversity. In wildland ecosystems, herbicides are sometimes used to reduce the density of exotic or invasive weed species and thus indirectly increase the diversity of native or desirable species. Where herbicides inadvertently enter the environment as contaminants, detrimental impacts on nontarget plants and other organisms may occur. However, the threat to plant biodiversity caused by habitat loss and the spread of

Encyclopedia of Biodiversity, Volume 3

exotic species is far greater than that caused by all forms of environmental pollution, including herbicides. This article reviews concepts of weeds; principles of weed management; and categories, action, and fate of herbicides. Impacts of herbicides on biodiversity of both target (weeds) and nontarget species and the role of weed control in preserving biodiversity are also discussed.

I. THE CONCEPT OF A WEED

Weeds are most often defined in human or anthropomorphic terms, that is, as plants growing where they are not wanted or plants out of place. More useful definitions of weeds are those that describe biological traits or characteristics. A list of "ideal characteristics of weeds," developed by Baker (1974), is widely known and cited for its description of traits that confer weediness on plants that possess them (Table I). These traits include germination under a broad range of conditions and over an extended period; rapid growth and prolific reproduction by sexual and asexual means; flexible breeding systems, including self-pollination, cross-pollination, and nonspecialized pollinators; effective seed dispersal; plasticity and tolerance of a breadth of environmental conditions; and adaptations for competitiveness. Though no plant species could possess all of these traits, plants considered to be major weeds are likely to possess many of them. Traits alone do not determine whether a plant will be a weed, but Baker's list is a useful tool for categorizing and studying plants that interfere with human activities and, thus, are called weeds.

TABLE I

Characteristics of the "Ideal Weed"[a]

Germination requirements fulfilled in many environments
Discontinuous germination (internally controlled) and great longevity of seed
Rapid growth through vegetative phase to flowering
Continuous seed production for as long as growing conditions permit
Self-compatible but not completely autogamous or apomictic
When cross-pollinated, unspecialized visitors or wind utilized
Very high seed output in favorable environmental circumstances
Produces some seed in wide range of environmental conditions; tolerant and plastic
Has adaptations for short- and long-distance dispersal
If a perennial, has vigorous vegetative reproduction or regeneration from fragments
If a perennial, has brittleness, so not easily drawn from ground
Has ability to compete interspecifically by special means (rosette, choking growth, allelochemicals)

[a] Source: Baker (1974). Reproduced with permission from the *Annual Review of Ecology and Systematics*, Volume 5, © 1974, by Annual Reviews.

A. Impacts of Weeds

By their very definition, weeds affect many activities in which humans are engaged. In agroecosystems their primary effect is to reduce crop yield and quality through competition for limited resources. Weeds also increase the time and costs required for crop production and interfere with harvesting. Indirectly, weeds affect crops through both positive and negative interactions with insect herbivores and their natural enemies. In rangelands, weeds possessing thorns or barbs pose physical hazards to livestock, and those containing toxins may cause allergies or poisonings of animals or humans. Other negative impacts of weeds include obstructing visibility around roadways, serving as a fire hazard, impeding use of recreation areas, and blocking the free flow of water in waterways, irrigation canals, and drainage ditches. In nonagricultural ecosystems, weeds often comprise the first stage of plant succession on land where the native vegetation has been disturbed. With increasing movement of humans across continental boundaries, exotic (nonnative) weed species have invaded many wildland ecosystems where human activities have disrupted the growth of indigenous (native) species. As a result, biological diversity has been reduced in many wildland areas that interface with urban areas. Other impacts of weed invasions in wildlands include alteration of ecosystem processes, support of nonnative animals, fungi, or microbes, and hybridization with native species to alter gene pools.

B. Weed Science

Since the beginning of agriculture, humans have employed various tactics to remove weeds from land where other uses are desired. In the United States today, billions of dollars are spent annually for weed removal using chemical and mechanical means. With the discovery of synthetic organic herbicides in 1941, weed science developed as a formal scientific discipline. Weed scientists have been extremely successful during this century in developing techniques to remove weeds from agricultural and other ecosystems. Over the past 50 years, weed science has grown into a multidisciplinary field of study encompassing fundamental research along with applied aspects of weed suppression. Today it encompasses researchers from numerous scientific disciplines, including chemistry, ecology, genetics, morphology, and physiology, as well as applied scientists and practitioners who focus on weed suppression. In recent years, increased awareness of environmental concerns has shifted the emphasis in weed science from a primary focus on herbicides to more integrated, ecological approaches for dealing with weeds.

II. WEED MANAGEMENT

In crop, forest, and rangeland production systems, as well as in wildlands, weed presence must usually be minimized to achieve a desired land use goal. Various tools and methods are used to suppress or remove weeds while not injuring the crop or desirable species. Weed management is a strategy that includes growing or fostering desired or beneficial vegetation while suppressing unwanted plants. For such management to be successful, knowledge of the biology, ecology, life history, and taxonomy of the weed species is required, as well as selection of the proper tools to use for their suppression.

A. Tools and Methods of Weed Management

The strategy of weed management includes three key components: prevention, eradication, and control. Prevention is keeping a weed from being introduced into an area where it does not already occur. Common preventive measures include using sanitary practices, eliminating weed spread through seeds and vegetative propagules, using quarantines, and following federal and state weed laws and regulations. Eradication is the total elimination from a particular area of a weed species and any plant parts capable of reproducing. Although eradication is often a stated goal of weed management programs, it is seldom achieved owing to the presence of seed banks and vegetative bud reserves in the soil of weed-infested areas. In contrast to eradication, weed control is the suppression or reduction of a weed species to an economically acceptable level. Under weed control programs, complete elimination of the weed is not the goal; instead, weeds are reduced to a level at which the cost of continued suppression does not exceed the value of the land or crop growing on it plus the benefit afforded by weed control. Once a weed is established in an area, weed control is the approach most commonly used to manage vegetation in that area over the long term. Methods used for weed control include biological, chemical, cultural, and mechanical techniques.

B. Integrated Approaches to Weed Management

With the growing recognition and concerns about the environmental impacts of agricultural practices, particularly the use of herbicides, integrated approaches for weed management have become commonplace. Integrated weed management (IWM) refers to a strategy for weed suppression that combines information on the biology and ecology of the weed with all available control technologies. Using this strategy, a variety of different methods are used in weed control, including nonchemical ones, as well as preventive measures, such that emphasis on herbicides is minimized. An alternative approach to crop production that is currently receiving widespread attention is sustainable agriculture, which refers to production systems in which external inputs, including synthetic fertilizers and pesticides, are minimized or avoided. However, current knowledge about the biology and ecology of weeds, their interactions with crops and wild plants, and nonchemical methods for their control is still limited. Thus, in most agricultural systems today, achieving weed control without herbicides requires extensive mechanical and hand labor, which is very costly. Until viable and economical alternatives to herbicides are available, most crop production and land management systems in the United States will continue to depend on some level of herbicide use for weed control.

C. Chemical Weed Control

Herbicides are chemicals used to suppress or kill plants or to interrupt their normal growth processes. Of all groups of pesticides (including insecticides, fungicides, and rodenticides), herbicides are the leading group in terms of tons produced, dollar value from sales, and total acreage treated. Extensive and widespread use of herbicides in agriculture continues because of their high level of effectiveness and low cost relative to other methods of weed control. Use of herbicides has resulted in improved control of weeds that grow within crop rows where cultivation is not possible. Herbicides have replaced frequent tillage operations in some systems, which conserves energy, reduces crop damage, and minimizes damage to soil structure. With herbicides, crop production is less dependent on weather and human labor, such that greater flexibility in choice of crops and management methods is possible. In the United States today, the abundance of relatively inexpensive food and fiber is due in large part to the benefits afforded by herbicides for weed control in the last half century. However, the use of herbicides also carries risks, including injury to crops and nontarget plants, herbicide residues in soil or water, toxicity to nontarget organisms, and concerns for human health and safety. For this reason, the benefits and risks of each method must be weighed carefully when developing a weed control program, particularly in wildland ecosystems. In the United States, all pesticide development and use is subject to strict regulation by the federal government.

1. Herbicides

There are approximately 150 herbicide active ingredients, which are formulated into hundreds of commercial products. Most are organic compounds, containing carbon, hydrogen, oxygen, and various other chemical elements. Each herbicide has a chemical name that describes its structure and a common name, which is often a simplified version of the chemical name. Formulated herbicides also have a trade name assigned by the manufacturer for marketing purposes. Manufacturers formulate herbicides to enhance their handling and weed control properties. Formulated compounds include the herbicide active ingredient plus inert ingredients such as solvents, diluents, and various additives. When the same herbicide active ingredient is formulated in more than one way, each is assigned a different trade name. Herbicides can be classified in several different ways, which provide users with a convenient means of selecting herbicides for various purposes.

Herbicides are often classified according to similarities in chemical structure, which often, but not always, result in similar effects on plants. A more useful classification scheme is based on where they may be used. In agriculture, herbicides are registered for use in agronomic and horticultural crops, turfgrass, and landscape and ornamental plantings. In noncrop areas, herbicides are used in pastures and rangelands, aquatic habitats, rights-of-way, utility sites, recreation areas, forests, and wildlands. Herbicides are also classified according to method and timing of application. Application methods include soil and foliage treatments, depending on where in the plant the chemical is most readily absorbed. Timings of herbicide application include preplant (prior to crop planting), preemergence (prior to crop or weed emergence), and postemergence (after crop or weed emergence). A herbicide classification scheme developed by Ross and Lembi (1999) is summarized in Table II.

Other classification schemes are based on plant responses to herbicides. Selective herbicides are more toxic to some plant species than others (e.g., monocotyledonous versus dicotyledonous plants), whereas nonselective herbicides are toxic to all plant species. Selectivity is one of the most important and useful characteristics of herbicides, for it allows applications to be made to weeds without risk of injury to crops or desirable vegetation. Factors related to herbicide chemistry (structure, formulation), the plant (age, size, surface characteristics, morphology, physiology), and the environment (humidity, temperature, soil moisture) determine the selectivity of a particular herbicide. Other categories of plant response used to classify herbicides are the pathway of herbicide movement in plants and the mechanisms by which herbicides kill plants (see Table II). Systemic herbicides are those that move in plants; movement occurs in the phloem (symplast), in the xylem (apoplast), or both. Herbicides that do not move in plants, but rather exert their effect at the site of application, are called contact herbicides. Herbicides classified by mechanism of action include growth regulators; inhibitors of photosynthesis, pigments, lipid synthesis, cell wall synthesis, amino acid synthesis, and cell division; and cell membrane destroyers.

2. Fate of Herbicides in the Environment

Herbicide fate in the environment is an issue of public concern and an important consideration when herbicides are registered for legal use. For herbicides to be effective, they must persist long enough to kill the weeds for which they were intended. Persistence beyond that time, however, may result in injury to nontarget plants and other organisms, residues in crops, and environmental contamination.

Herbicides that enter plants generally move to a site of action and cause a toxic reaction. Over time, most herbicides in plants are transformed into relatively less toxic forms by biochemical processes. If a herbicide is not degraded, it may remain in the plant or end up in the soil as a contaminant. Eventually, however, all herbicides that enter a plant, the soil, water, or atmosphere will be degraded by the same chemical and physical reactions that act on biologically derived compounds.

Once a herbicide reaches the soil, several processes, including adsorption to soil particles, movement to another location, and decomposition, will determine its persistence. Soil and herbicide characteristics regulate adsorption of herbicide molecules onto particles of clay and organic matter. Herbicides that are tightly adsorbed are not available for plant uptake, movement to other

TABLE II

Herbicide Classification According to Method of Application, Movement in Plants, and Mode of Action (Examples of Herbicide Classes or Individual Herbicides in Each Category Are Listed)[a,b]

Method of application	Movement in plants	Mode of action	Chemical class or individual herbicide
Foliar	Translocated in phloem	Auxin-type growth regulators	Phenoxy acid herbicides, benzoic acid herbicides, picolinic acid herbicides, naptalam
		Aromatic amino acid (EPSPS) inhibitors	Glyphosate
		Branched-chain amino acid (ALS/AHAS) inhibitors	Sulfonylurea herbicides, imidazolinone herbicides, triazolopyrimidine sulfonanilide herbicides, pyrimidinyl oxybenzoate herbicides
		Carotenoid pigment inhibitors	Amitrole, clomazone, fluridone, isoxaflutole, norflurazon
		Lipid biosynthesis (ACCase) inhibitors	Aryloxyphenoxy propionate herbicides, cyclohexanedione herbicides
		Organic arsenicals	DSMA, MSMA
		Unclassified herbicides	Asulam, difenzoquat, fosamine, propanil
Foliar	Translocated in xylem	Photosystem II (PSII) photosynthetic inhibitors	s-Triazine herbicides, metribuzin, phenylurea herbicides, uracil herbicides
		Other photosynthetic inhibitors	Bentazon, bromoxynil, phenylcarbamate herbicides, pyrazon, pyridate
Foliar	Contact (not translocated)	Photosystem I (PSI) cell membrane destroyers	Bipyridilium herbicides
		Protoporphyrinogen oxidase inhibitors	Diphenylether herbicides, oxadiazole herbicides, flumiclorac, triazolinone herbicides
		Glutamine synthesis inhibitors	Glufosinate
Soil	Depends on herbicide	Microtubule/spindle apparatus inhibitors (root inhibitors)	Dinitroaniline herbicides, DCPA, dithiopyr, pronamid
		Shoot inhibitors	Chloroacetamide herbicides, thiocarbamate herbicides
		Miscellaneous cell division inhibitors	Bensulide, napropamide, siduron
		Cell wall formation inhibitors	Dichlobenil, isoxaben, quinclorac

[a] Abbreviations: EPSPS, 5-enolpyruvylshikimate-3-phosphate synthase; ALS, acetolactate synthase; AHAS, acetohydroxy acid synthase; ACCase, acetyl-coenzyme A carboxylase; DSMA, disodium methanearsonate; MSMA, monosodium methanearsonate; DCPA, dimethyl 2,3,5,6-tetrachloro-1,4-benzenedicarboxylate.

[b] Source: Adapted from Ross and Lembi (1999).

sites, or decomposition. Herbicides that are loosely adsorbed or located in the soil solution may move by leaching (vertical and lateral movement by water). The potential for movement of herbicides into groundwater through leaching is a concern for a few very mobile herbicides, which consequently are subject to strict monitoring and regulation. Herbicides may also enter the environment during application by drift (movement of herbicide particles in air) or from the soil surface by volatilization (conversion into a vapor). Both of these processes can be minimized or avoided when proper equipment and application techniques are used.

Herbicide decomposition occurs in soil, air, water, plants, animals, and microorganisms and results in breakdown of the original herbicide molecule and loss of herbicide activity. Decomposition of herbicides occurs by photochemical (breakdown in sunlight), chemical, or microbiological means. Herbicide decomposition also depends on temperature and other environmental factors, plus the concentration of herbicide applied. Products of herbicide decomposition may eventually degrade into simple organic molecules. Because most herbicides are degradable, they do not build up in the soil over time, even after repeated use. Additionally, the soil microbial populations are adaptable such that the application of a pesticide to soil is often followed by an increase in the number of microbes that can degrade it. When processes that regulate herbicide fate are understood, and the legal requirements for herbicide application and use are followed, minimal contamination of the environment should result.

3. Regulation of Herbicide Use

All pesticides, including herbicides, must be registered with the U.S. Environmental Protection Agency (EPA) before they can be distributed or sold in the United States. Two laws, the Federal Insecticide, Fungicide and Rodenticide Act (FIFRA), and parts of the Food, Drug, and Cosmetic Act (FDCA) regulate pesticide development and use. FIFRA provides for registration and cancellation of pesticides, maintains a classification system for pesticides based on toxicity, and allows states to regulate pesticide use in a manner consistent with federal regulations. The FDCA mandates establishment of tolerances for pesticides in food, feed, fiber, and water. These laws were written to ensure that benefits from the use of pesticides are in balance with concerns about health and environmental impacts. Each state also has laws regulating pesticide use, including worker safety regulations and requirements for use of the most toxic (restricted use) pesticides.

For a pesticide to be registered by the EPA, it must be subjected to over 100 safety and environmental tests. Information required before registration includes chemical and physical properties, environmental fate, amounts of the pesticide in feed and food crops, toxicological properties, and effects on nontarget plants and animals. Such data are usually required both for the pesticide and for its metabolites, or breakdown products. With this information, the relative benefits and risks of each pesticide can be determined.

Pesticide effects on human health are expressed as toxicity, the amount of the chemical that is harmful or lethal, and exposure, the probability of encountering a harmful dose of the chemical. The combination of acute toxicity plus exposure to a pesticide during its expected use determines the hazard it poses to humans. The EPA uses these data to set a tolerance level for each agricultural pesticide, which is the maximum amount of the chemical allowed on a particular crop. These and many other data, including chronic toxicity, reproductive effects, teratogenicity, and carcinogenicity, are used by the EPA in setting limits for pesticide use to ensure that hazard to humans from use of pesticides will be at acceptably low levels. Toxicology and exposure studies are also required on certain species of wildlife, including birds, fish, and invertebrates, before pesticides can be registered. These tests utilize studies of pesticide residues in foods that these species may consume, as well as potential concentrations of pesticide in water or air to determine hazard to nontarget species from pesticide use.

Pesticides vary widely in their toxicological properties. In relation to all pesticides, which includes insecticides, fungicides, and rodenticides, most herbicides are relatively nontoxic to mammals, since many of the processes or pathways they inhibit in plants are not present in mammalian systems. For the purposes of regulation, pesticides are classified according to their toxicity, which is a relative term used to describe the amount of a chemical that causes harm to a particular species. The most common unit of measurement for toxicity is the lethal dose, LD_{50}, or lethal concentration, LC_{50}, which is the dose or concentration that kills 50% of the test population, respectively. High values for LD_{50} or LC_{50} indicate lower toxicity, as higher doses are required to produce lethal effects. The categories of acute toxicity that must be shown on herbicide labels are listed in Table III. Because all herbicides are toxic to some degree, container labels are required to give specific directions for use as well as ingredients, properties, hazards, exposure limits, first aid procedures, and other information. When handled according to the label directions, hazard to humans and wildlife from exposure to herbicides can be avoided.

III. EFFECTS OF HERBICIDES ON BIODIVERSITY

A. Weeds

All weed control practices exert selection pressure on weeds and thus can have short- and long-term effects

TABLE III

Toxicity Categories, Lethal Dosages, and Signal Words Used in Herbicide Labeling

Toxicity category	Toxicity	Signal word	Oral LD_{50} (mg/kg)	Dermal LD_{50} (mg/kg)
I	Very high	Danger	≤50	≤200
II	High	Warning	51–500	201–2000
III	Moderate	Caution	501–5000	2001–20,000
IV	Low	Caution	>5000	>20,000

on the composition, structure, and dynamics of weed communities. Agricultural weed communities tend to have lower species diversity than natural plant communities and are often dominated by a few key species. The primary short-term effect of weed control is a reduction in weed density, particularly of the dominant species, which is desirable in order to improve crop yields or facilitate land use. Over a longer time frame, weed control practices rarely eliminate weeds altogether, rather they generally result in changes in species composition and structure of weed communities. In the case of herbicides that act on specific plant processes, selection pressure over time can eliminate susceptible genotypes and thus cause evolutionary changes in weed populations. Most research in weed science to date has emphasized reducing weed density and improving crop yields. Only recently has attention been focused on changes in weed community dynamics as a result of weed control practices. Thus, only a few generalities can be made about specific effects or directions of change in weed diversity caused by biological, cultural, and mechanical methods of control. The most information available on the role of weed control in shaping weed communities comes from documented cases of herbicide-resistant weeds, which have been selected by repeated use of the same herbicide or herbicide class. It is clear that herbicides are a powerful evolutionary force acting on weed communities. Nevertheless, the data that are available indicate that effects of nonchemical forms of weed control on weed community dynamics are also significant and warrant further study.

1. Effects of Weed Control Practices

Changes in weed community composition and structure due to various agricultural practices have been documented; however, assessments of the effects of these practices on weed species diversity have been made only rarely. In a review of integrated weed management, Clements *et al.* (1994) calculated diversity indices from an array of published data to compare the impacts of conventional and alternative weed management practices on weed species diversity. When compared to weed control by mechanical cultivation only, broadcast applications of herbicides resulted in lower weed species diversity over time. Where herbicide use was reduced by placement of applications only in bands over the crop row, higher diversity of weeds resulted than when broadcast herbicide applications were made. Although these findings suggest that use of herbicides reduces weed species diversity, actual case studies show that this generalization is too simplistic and that impacts on diversity depend on the persistence of the specific herbicide used. When different herbicides were evaluated, applications of preemergence herbicides (those having residual soil activity) reduced weed species diversity more than applications of postemergence herbicides (those with no residual activity). By exerting continuous selection pressure on susceptible species from the time of emergence, preemergence herbicides often reduce the richness and diversity of weed communities. In contrast, postemergence herbicides are present in the environment only after emergence and for a shorter time period, which may permit a more diverse weed flora to establish both before and after the disturbance of herbicide application is imposed.

Even in cases where weed species diversity is relatively unaffected by herbicide applications, interspecific selectivity of many herbicides causes a shift within a weed community from species that are susceptible to species that are naturally more tolerant to the particular herbicide. A common example of this phenomenon is the shift in relative abundance from dicotyledonous (dicot) weed species to monocotyledonous (monocot, usually grass) weed species following repeated use of the herbicide 2,4-D for control of dicot weeds in cereal (grass) crops. Similarly, in fields where herbicides have been used over many years for control of annual weed species, shifts in the weed flora to predominantly perennial weed species are commonly observed. Weed species compositional shifts also occur when the weeds in a field are taxonomically related to the crop species grown there, since plants often respond similarly to herbicides when they are in the same taxonomic family. Thus, many cases have been documented where repeated use of a particular herbicide in a crop has selected for weeds that are in the same plant family as the crop. Despite the increasing documentation of changes in composition of weed communities as a result of herbicide use, only recently has attention been focused specifically on weed species diversity.

Integrated and alternative weed management methods, which employ a combination of tools to control weeds below a specified threshold, theoretically should not impose strong directional selection on weed populations. Thus, weed species diversity might be expected to increase, or at least not decrease to the same extent as under a regime of chemical weed control, when a variety of tools are used. Clements *et al.* (1994) point out the many questions that remain to be answered about the potential effects of integrated weed management techniques on diversity of weeds. In addition, the role and potential importance of weed biodiversity in agroecosystems remain to be defined.

2. Selection of Herbicide Resistance

Herbicide resistance represents an extreme shift in weed species composition caused by the selection of plants possessing a gene or genes for resistance to a particular herbicide within a species that was formerly susceptible. In cases where a particular herbicide has been used repeatedly over several years, resistant weeds may be selected and come to dominate a weed community such that species diversity declines. Since the 1970s, many cases of the evolution of resistance have been documented as a result of repeated herbicide applications for weed control. Table IV summarizes the worldwide occurrence of resistant weed biotypes to different herbicide groups.

Several characteristics of herbicides and their use contribute to a high probability for selection of resistance in weeds. These include having a single target site and specific mechanism of action, being extremely active and effective in killing a wide range of weed species, and having long soil residual activity and season-long control of germinating weeds. In addition, frequent application of a particular herbicide over several growing seasons without rotating, alternating, or combining with other types of herbicides contributes to a high risk for evolution of resistance. Some herbicides are thought to pose a low risk for selection of resistance owing to their nonspecific mechanism of action and short or no soil residual activity. Even in these cases, however, repeated use of the same herbicide will exert selection pressure on weeds. Recommendations for preventing or managing herbicide-resistant weeds include practices such as rotating herbicides from different chemical classes to avoid imposing the same selection pressure over time and integrating a combination of weed control methods. The higher level of weed species diversity that presumably would result from these approaches should reduce the potential for propagation of herbicide-resistant genes in weed populations (Clements *et al.*, 1994).

TABLE IV

Worldwide Occurrence of Resistant Weed Biotypes to Different Herbicide Classes[a,b]

	Resistant weed biotypes			
Herbicide class	Dicot species	Monocot species	Total	Number of countries
Triazines	42	19	61	22
ALS inhibitors[a]	43	20	63	18
Bipyridiliums	18	7	25	13
Ureas/amides	6	11	17	19
Synthetic auxins	15	4	19	12
ACCase inhibitors[a]	0	21	21	18
Dinitroanilines	2	7	9	5
Triazoles	1	3	4	2
Chloroacetamides	0	3	3	4
Thiocarbamates	0	3	3	3
Nitriles	1	0	1	1
Glycines	0	2	2	3
Benzoflurans	0	1	1	1
Organoarsenicals	1	0	1	1
Carbonic acids	0	1	1	1
Pyrazoliums	0	1	1	2
Totals	129	103	232	

[a] ALS, acetolactate synthase; ACCase, acetyl-coenzyme A carboxylase.

[b] Source: International Survey of Herbicide-Resistant Weeds, compiled by I. Heap and available at www.weedscience.con.

B. Other Organisms

By suppressing, removing, or destroying vegetation, weed control modifies the environment and habitat of other organisms. In agroecosystems, crop pests as well as beneficial organisms can be affected by weed removal since weeds can serve as host plants or food sources for many types of organisms, including insects, fungi, and nematodes. In some cases, a high diversity of weed species in an agricultural field has been shown to reduce the magnitude of insect attacks on crop plants because the weeds serve as alternate food for the insect pests or harbor beneficial organisms that feed on the pests. In other cases, however, increased weed species diversity results in increased insect problems in a crop field because the weeds provide a food source or habitat so that the insect pest can remain in the field even during periods when the crop is absent. In those cases, weed control to reduce the diversity of weeds in a particular field will decrease the incidence of insect pest damage to the crop. To the extent that herbicides affect weed biodiversity, therefore, their use will also indirectly affect insect organisms.

The response of fungal and nematode populations to weed species diversity has not been well studied. In general, weed control practices result in a cleaner crop field, which usually leads to fewer disease and nematode problems. However, it is also possible that large areas of crop monocultures with few weeds may be susceptible to widespread disease epidemics because of the lack of genetic diversity in response to the disease. To date, little information is available to indicate how weed biodiversity or use of herbicides in weed control influences populations of fungal pathogens or nematode pests of crop plants.

The question of whether weed biodiversity is an asset or a detriment to overall pest management in agriculture deserves serious attention by researchers. To answer this question, more information is needed on the effects of various weed control practices both on weed biodiversity and on nonweed organisms. With this information, the costs and benefits of weed control can be weighed against potential costs and benefits of alternative strategies to control other pests that are affected by weed control. The question is complicated by the fact that in large-scale, mechanized agricultural production, increases in weed species diversity complicate weed control efforts. However, this problem could be offset if there were benefits to be gained from maintaining increased genetic diversity of weeds in a field, such as greater buffering against selection for herbicide resistance.

In wildland situations where exotic weeds have replaced native vegetation, weed control is increasingly practiced to reduce weed invasions and restore the abundance and diversity of native plants. The indirect result of these weed control activities is often a restoration of habitat for nonplant organisms, such as birds and small mammals. In these situations, therefore, weed control, including use of herbicides, can result in increased diversity of other organisms that depend on native plant communities for habitat.

C. Wild Plants

Weeds are managed in many situations to restore or preserve the biodiversity of native plant species. As described in Randall (1996), the most common methods used for weed management in wildlands in the United States are manual and mechanical weed removal, prescribed fire, release of biological control agents, use of grazing animals, encouragement of native competitors, and judicious use of herbicides. Weed management strategies in wildlands differ from those in agroecosystems because wildland managers must promote or protect large numbers of plant and nonplant species rather than one or a few crop species. Thus, wildland weed management is generally approached with a desire to manage an entire plant community rather than to control a single weed species. Although the same methods can be used in agricultural and nonagricultural habitats, wildland managers must minimize negative impacts to a wider range of nontarget species than must agricultural land managers. Therefore, the methods used for wildland weed control are often labor-intensive and more environmentally conservative than in agroecosystems.

All weed control methods incur some risk to the environment in which they are used, which must be weighed against the risk of taking no action and allowing weeds to continue to spread. Mechanical weed control disturbs the soil, destroys vegetation, and leaves gaps that may be reinfested with weeds. Biological control agents may attack nontarget species or become adapted and spread in undesirable ways. Herbicides may unintentionally kill nontarget plant species, indirectly impacting habitat for other organisms, and may become environmental contaminants if not used properly. In most cases, the herbicides used in wildland situations are postemergence compounds with very short or no soil residual activity to minimize their effect on nontarget plant species. However, few empirical data exist on either intentional or unintentional effects of herbicides on native plant communities.

Rice *et al.* (1997) conducted a long-term field study on the effects of herbicides used for the control of an exotic species on the structure and species diversity of native plant communities. Herbicide treatments were highly effective in controlling the exotic weed (*Centaurea maculosa* Lam., spotted knapweed) and converting the plant community back to its native composition. Overall, only small, short-term depressions in species richness and diversity resulted from herbicide use. As shown in agroecosystems and as discussed earlier, however, the rate (dosage), frequency, and timing of herbicide applications are important determinants of plant community responses. Higher rates, more frequent applications, or applications earlier in the growing season have been shown to reduce the diversity of nontarget species more than lower rates, less frequent applications, or later applications. Similarly, selectivity of the herbicide as well as precision of the application method can be managed to reduce impacts on nontarget species. Rice *et al.* (1997) concluded that periodic application of appropriate herbicides can be used to restore native plant communities from dominance by exotic species and to maintain their diversity as well. Transient reductions in diversity that may result from herbicide use are likely to be negligible when compared to the serious impacts on native communities created by exotic plant species.

As in agroecosystems, questions remain in natural systems about the role and management of weed abundance and diversity in wildland ecosystems. Where exotic weed invasions are widespread, weed removal alone may not result in establishment of native species without additional inputs such as revegetation with desirable species. In some areas, invasive exotic weeds have replaced native species and become an important source of food and cover for native birds and mammals, which must be considered before the weeds are removed. The use of herbicides in wildlands is often controversial because of the risks of environmental contamination, as well as the general public perception that herbicides are dangerous. Just as in agriculture, therefore, the costs and benefits of weed control in wildlands must be weighed before any method is chosen. In cases where herbicides are used, the determination is made that the risks of herbicide use are more than offset by the benefits gained in increased or restored diversity of native species due to control of exotic weeds.

IV. CONCLUSIONS

Herbicides have been employed for decades to control unwanted vegetation in agroecosystems, yet their impacts on weed biodiversity and the implications of those impacts for weed management have not been well studied except in the case of the evolution of herbicide resistance. Increasing emphasis over the past decade on the importance of maintaining diversity of all genetic resources has rarely been extended to agroecosystems, despite the likelihood that assessments of the nature and importance of weed biodiversity might reveal an important role for weeds in agroecosystem stability and sustainability. Based on experiences with herbicide use in agriculture, deliberate use of herbicides to control invasive weed species in wildlands has become increasingly common, yet the impacts of herbicides on the composition, structure, and dynamics of nontarget native plant communities are poorly understood and usually assumed to be negative. When placed in a comparative context, however, habitat degradation and destruction, particularly by competition with exotic species, is a much more pervasive threat to biodiversity of endangered plant species than is pollution of all forms, including agricultural pesticides (Wilcove *et al.*, 1998). Thus, it is imperative that research be continued into the effects of all forms of weed control on biodiversity of both target and nontarget species.

See Also the Following Articles

AGRICULTURE, INDUSTRIALIZED • ECOLOGY OF AGRICULTURE • HERBACEOUS VEGETATION, SPECIES RICHNESS IN • INSECTICIDE RESISTANCE • PESTICIDES, USE AND EFFECTS OF

Bibliography

Ahrens, W. H. (ed.). (1998). *Herbicide Handbook*, 8th ed. Weed Science Society of America, Champaign, Illinois.

Baker, H. G. (1974). The evolution of weeds. *Annu. Rev. Ecol. Systematics* 5, 1–24.

Clements, D. R., Weise, S. F., and Swanton, C. J. (1994). Integrated weed management and weed species diversity. *Phytoprotection* 75, 1–18.

Derksen, D. A., Thomas, A. G., Lafond, G. P., Loeppky, H. A., and Swanton, C. J. (1995). Impact of post-emergence herbicides on weed community diversity within conservation-tillage systems. *Weed Research* 35, 311–320.

Devine, M. D., Duke, S. O., and Fedtke, C. (1993). *Physiology of Herbicide Action*. Prentice-Hall, Englewood Cliffs, New Jersey.

Dyer, W. E., Hess, F. D., Holt, J. S., and Duke, S. O. (1993). Potential benefits and risks of herbicide-resistant crops produced by biotechnology. *Horticult. Rev.* 15, 367–407.

Froud-Williams, R. J. (1988). Changes in weed flora with different tillage and agronomic management systems. In *Weed Management in Agroecosystems: Ecological Approaches* (M. A. Altieri, and M. Liebman, eds.), Chap. 13. CRC Press, Boca Raton, Florida.

Haas, H., and Streibig, J. C. (1982). Changing patterns of weed distribution as a result of herbicide use and other agronomic factors. In *Herbicide Resistance in Plants* (H. M. LeBaron, and J. Gressel, eds.), Chap. 4. John Wiley & Sons New York.

Heap, I. (1998). *International Survey of Herbicide Resistant Weeds.* Website on the Internet: www.weedscience.com.

Kirkwood, R. C. (chair). (1997). Biodiversity and Conservation in Agriculture. BCPC Symposium Proceedings No. 69. British Crop Protection Council, Surrey, United Kingdom.

Norris, R. F., and Kogan, M. (2000). Interactions between weeds, arthropod pests, and their natural enemies in managed ecosystems. *Weed Science* 48, 94–158.

Powles, S. B., and Holtum, J. A. M. (1994). *Herbicide Resistance in Plants. Biology and Biochemistry*. Lewis Publishers, Boca Raton, Florida.

Prather, T. S., DiTomaso, J. M., and Holt, J. S. (2000). *Herbicide Resistance: Definition and Management Strategies*. Available on the Univ. of Calif. Division of Agriculture and Natural Resources Website on the Internet: anrcatalog.ucdavis.edu/pdf/8012.pdf.

Radosevich, S. R., Holt, J. S., and Ghersa, C. M. (1997). *Weed Ecology. Implications for Management* 2nd ed. John Wiley & Sons, New York.

Randall, J. M. (1996). Weed control for the preservation of biological diversity. *Weed Technol.* 10, 370–383.

Rice, P. M., Toney, J. C., Bedunah, D. J., and Carlson, C. E. (1997). Plant community diversity and growth form responses to herbicide applications for control of Centaurea maculosa. *J. Appl. Ecol.* 34, 1397–1412.

Ross, M. A., and Lembi, C. A. (1999). *Applied Weed Science*, 2nd ed. Prentice–Hall, Upper Saddle River, New Jersey.

Stevenson, F. C., Légère, A., Simard, R. R., Angers, D. A., Pageau, D., and Lafond, J. (1997). Weed species diversity in spring barley varies with crop rotation and tillage, but not with nutrient source. *Weed Science* 45, 798–806.

Wilcove, D. S., Rothstein, D., Dubow, J., Phillips, A., and Losos, E. (1998). Quantifying threats to imperiled species in the United States. *BioScience* 48, 607–615.

HIGH-TEMPERATURE ECOSYSTEMS

Richard G. Wiegert
University of Georgia

GLOSSARY

acidophilic Organisms preferring or requiring a low pH environment.

chemoautotrophs Organisms able to synthesize organic compounds by the oxidation of energy-rich inorganic sources. Light is not used.

cyanobacteria A group of bacteria containing chlorophyll and capable of photosynthesis.

eucaryotes Organisms possessing a defined cell nucleus and nuclear membrane. Includes all organisms above the level of the procaryotic bacteria.

frustule The hard, silica-containing skeleton of diatoms (green algae).

geothermal heating Water heated at depth in the earth and released to the surface as thermal outflows.

heterotrophs Organisms dependent for their energy on organic compounds. If oxidized, the process is aerobic. When oxygen is absent, the process is anaerobic (fermentation).

metazoa Organisms whose individuals have more than one cell (multicellular).

photoautotrophs Organisms able to synthesize organic compounds from water and inorganic nutrients, using the energy in photons of light.

procaryotes Organisms without a defined cell nucleus or nuclear membrane.

protozoa A group of eucaryotic organisms usually classified as animals and regarded as primitive. But some major problems exist with some protozoa such as slime molds and euglenoid forms, which share many of the properties of plants.

solar heated Water raised to temperatures significantly exceeding the regional temperatures of lakes and streams. Many of the thermally tolerant and thermal-opportunist groups are thought to have evolved in shallow solar-heated ponds and water margins.

thermal opportunist Organisms that have developed life history characteristics that permit the exploitation of temporary cooler spots in thermal systems.

thermal systems Outflows of geothermally heated water, usually thought of in terms of those at the surface of the earth (see *thermal vent*) where light is present.

thermal vent Geothermally heated water issuing from cracks in the ocean floor. Very high temperatures and pressures and the complete absence of light are characteristic.

thermally tolerant Organisms able to live at temperatures significantly higher than the regional norm. They can also exist at cooler temperatures, but may not compete well in the latter situation.

thermophily Often restricted to mean those organisms adapted to living at high temperatures and unable to survive at lower temperatures. In this chapter the term in is also used in a more general way to characterize any type of adaptation to thermal systems.

THIS CHAPTER ON THERMAL BIODIVERSITY considers the number and kinds of organisms that are found in the thermal systems and thermal vent communities of the world.

I. EXTREME ENVIRONMENTS

The diversity of organisms in thermal environments is determined, in common with all biodiversity, by the interactions of the physical/biotic environment with the adaptive abilities of individual species populations. But in some respects thermal environments (ecosystems) are quite different, even unique, compared to ecosystems dominated by weather, geological origins of soils, and patterns of precipitation and drainage. First of all, thermal ecosystems form a class of extreme environments characterized by a mean temperature markedly higher than "normal." (Other extreme environments possess one more environmental variable far outside the "norm"—for example, arctic, desert, brine lakes, low pH systems, etc.) Some of these, such as the arctic/antarctic frozen wastes and extreme deserts, are caused by widespread climatic factors and are typically very large, whereas thermal ecosystems and some other extreme environments are small and result from local changes in the physical environment.

What is "normal" and what effect does the extreme factor or factors have on biodiversity? Here I define an extreme environment as "a condition under which some organisms can grow, whereas others cannot." This is too general for the purpose of this chapter, because each different ecosystem has its adapted flora and fauna. A more narrow definition identifies the extreme as excluding entire higher taxonomic groups, not simply species. Thus the extreme environment is not only low in the diversity of species, but also low in the diversity of higher orders of classification as well.

II. THERMAL SYSTEMS—GENERAL CHARACTERISTICS AND DEFINITIONS

Among extreme environments, thermal ecosystems are defined as those with an elevated temperature, compared with the range of other ecosystems found at the same latitude and elevation. The source of the heat may be (a) solar energy, (b) anthropogenic activities, or (c) geothermal. The organisms inhabiting these heated ecosystems can be classified as thermophilic, thermal tolerant, or thermally opportunistic. The first of these, thermophiles (the general term used in this chapter), require the high temperature for survival to which they are adapted optimally. Thermally tolerant organisms have evolved mechanisms to enable them to exist at extreme temperatures, but they can also survive at lower temperatures. Indeed, in culture they might do better physiologically at these lower temperatures. Competition with other nonthermal species may, however, restrict their distribution in nature. Thermal opportunists, on the other hand, have evolved mechanisms to avoid the effects of direct immersion in hot water. They seek out and take advantage of temporary refugia by using rapid maturation times, rapid mobility, and short life histories. I will later discuss the evolution of these traits and the groups to which they apply.

Solar heating of vascular plant leaves and of the soil/litter surface is enhanced by dark coloration, abundance of sunlight, and the slope of the land surface. Temperatures of plant leaves in desert environments may reach 60°C; the same temperature applies to soil/litter systems, but it is commonly much lower, and the soil litter surface cools rapidly at night and during rains. Furthermore, the soil/litter surface (as well as plant leaves) heated by the sun is very thin because of its insulating value, and the system is very dry when heated. Thus motile organisms can move out of these thermal systems during periods of inhospitable temperature regimes. Vascular plants can avoid the effects of heat by being rooted in the cooler subsurface soil, and nonmotile microorganisms resist the thermal effects by forming spores or other heat-resistant quiescent stages. The net result is that there is little effect of solar heating on the biodiversity of these ecosystems compared to adjacent systems with more shading of the soil surface. There may, of course, be a loss of species in the driest of these systems, but as a consequence of lack of water, not of temperature per se. Solar heating may also act on very shallow aquatic systems; the margins of lakes and ponds in temperate and tropical latitudes are examples. This can provide a habitat for certain multicelled animals adapted to warm waters, but it does not exclude animals that simply avoid the shallows during the day and forage there at night. Temperature increases are seldom very great, and biodiversity in these systems is affected very little. A case involving a thermally adapted dragonfly will be discussed later in this chapter.

Anthropogenic activities include steam vents, hot water plumbing and heating systems, and fermentation. Where liquid water condenses around steam vents, certain thermal microorganisms can be found and some of them can also be found in hot water systems. But these are too small and transient to be of interest to the student of thermal biodiversity. Fermentation, however, is a large natural phenomenon, as well as the result of certain human activities, such as composting and piling up tailings from coal mines. As a natural process, the accumulation of fermenting organic matter is seldom large enough or accumulating fast enough for the heat of fermentation to significantly increase the temperature. Plant debris added to the soil litter system during the year is simply decomposed at a relatively constant rate and the heat produced is dissipated to the environment. Humans, however, often pile up organic wastes to a degree that the insulating value of the pile causes a rise in the internal temperature, killing, for example, the seeds of vascular plants in compost heaps and excluding all but microorganisms from the water draining from coal tailing piles. However, the latter effect is often caused more by the extremely low pH of such systems than from elevated temperatures. Although these environments are of little interest from the standpoint of biodiversity, they are of great interest to the student of microbial taxonomy. For example, an entire genus, *Thermoplasma*, is restricted to coal refuse piles.

III. GEOTHERMALLY HEATED SYSTEMS-CLASSIFICATION AND DESCRIPTION

The extreme ecosystems characterized by some form of geothermal heating of water are small, both in absolute size (some are only a few square meters or less in extent) and in relation to the surrounding regional climate and soil-dominated areas. Other extreme environments caused by aridity, extreme salinity, and low temperatures stress organisms by posing problems of getting or retaining water. In contrast, geothermal systems have water, indeed they are defined on the basis of hot water. The problems they pose for organisms attempting to survive and grow involve adaptations for tolerating or avoiding the hot water. Biotic diversity is typically low, both in absolute numbers of species and diversity of higher taxonomic groups. This diversity is also usually low relative to the environments surrounding the thermal systems, unless the latter are situated within a larger extreme environment. The deep-sea thermal vents are an example of the latter.

Most thermal ecosystems of the world are caused by the surface emergence of geothermally heated water. The mechanisms of heating involve the percolation of surface water to some depth, where it is heated and pushed again to the surface. The differences in thermal and nutrient characteristics of the emerging water depend on the substrate rocks it saturates after being heated and the length of time the water travels just under the surface before emerging in heated lakes, pools, streams, and spring seeps. Thermal systems are usefully catalogued into four types. These take the form of (a) heated pools or small lakes, (b) small streams, and (c) outflows from thermal springs. These first three types all exist under one atmosphere of pressure or less and the highest temperature found is 100°C, the boiling temperature at sea level. A fourth type (d) the thermal vent communities form when heated water is ejected at depth on the ocean floor, and temperatures in the hundreds of degrees centigrade are possible. These were first discovered in 1977; they are small, usually less than a few hundred square meters, and may vent water (black smokers) that reaches 450°C.

The remainder of this chapter focuses on the various manifestations of these heated ecosystems. All of them share the characteristic of being aquatic, but they show differences related to the volume and movement of water flowing out of the ground as well as the mechanisms available to organisms for adapting to the thermal environment. If temperature is regarded as analogous to a toxin, an organism can exist by (a) adapting to exclude the factor (in the case of temperature, this amounts to avoiding lethal levels for a particular stage in the life history), (b) adapting its structure/life history/physiology to nullify the effects of the factor, or (c) simply adapt in more minor ways to live with the factor. The second and third strategies differ quantitatively but not qualitatively. For example, enzymes can be changed to different temperature optima by relatively small evolutionary steps, but to cope with temperatures that will destroy (denature) the enzyme, a major new structure must be developed.

Geothermally heated lakes and pools characteristically are well mixed thermally, with little or no directional current. This produces, typically, a body of water with a relatively homogeneous temperature regime. The consequences for organisms attempting to colonize such ecosystems is that they cannot escape the thermal load, thus any organisms with limited motility must adapt to survival at the temperature regime of the system. Algal/bacterial mats in such thermal ecosystems, at temperatures where photosynthesis is possible (discussed later), form on the bottom of the lake or pool. There is little or no thermal gradient in these systems, so any eucaryotic consumers must adapt to the prevailing temperature regime or fail to colonize. Thus in Yellowstone National Park in the United States and in other geothermally active areas of the world, the thermal lakes and pools can be ranked from low to higher diversity as an inverse function of the mean maximum annual temperature of the system. Because of the lack of strong currents and replacement of surface water, cooling of these systems is relatively slow and the annual variation in temperature is low, even in the cold temperate latitude and high elevation of Yellowstone Park.

Thermal streams share with thermal lakes and pools the characteristic that the layer of productive organic photo- or chemoautotrophs is on the bottom and organisms attempting to colonize must adapt to tolerate the temperature regime. But the stream has a significant unidirectional flow and hot water emerging from the substrate onto the surface immediately begins to cool. The result is the establishment of a strong thermal gradient from the source downstream, within which different specific conditions for community development can be found. In theory, the cooling curve will be exponential, but other factors intervene, among which the most important are the current turbulence and the temperature, humidity, and the wind speed immediately above the surface of the stream. For example, in a small Yellowstone thermal stream (Gentian Stream, Firehole Lake Drive), I have found that the temperature in summer can fluctuate rapidly as the sun goes behind clouds and wind velocity changes. This effect takes place in minutes, whereas longer lasting weather changes (cloudy days, for example) cause changes in hours to days. In Yellowstone, 250 m from the source, the annual fluctuation in temperature is approximately 20°C, significant, but also a testimony to the thermal capacity of water since the flow of this small stream is measured in liters per minute. Thus in contrast to the organisms in thermal lakes and pools, the inhabitants of geothermally heated streams are far from living in a constant temperature natural chemostat, unless the stream habitat is very close to the source.

Outflows from thermal springs differ from lakes, pools, and streams in the volume of water discharged, its depth and pattern of flow after discharge, and in the types of communities and organisms that colonize the outflows. In contrast to lakes and pools, spring discharges have strong flows and turbulent mixing, with a relatively steep thermal gradient established from the source to the point where temperature of the water approximates that of ambient air. Outflows do not generally follow a well-defined channel (except near some of the sources), and in any case the depth is shallow. Once the temperature reaches the point that is tolerated by the filamentous cyanobacteria (blue green algae), a mat of microorganisms forms that can thicken and directly affect the flow of the water. This creates a community with distinct cool patches; some are cool on top with hot water flowing underneath, others are devoid of flow. This temperature heterogeneity provides an additional manner in which motile organisms can colonize thermal communities, by avoiding lethal temperatures as a consequence of adapting physiology, life history, or behavior to take advantage of the temporary cooling of the algal mat. Spring outflows exhibit considerable variety. The sources are of different temperatures, primarily a consequence of the water traveling for variable distances in the soil after emergence from the underlying rock and before emergence onto the surface. The chemical characteristics of the emerging waters may differ (although I am eliminating from this discussion those waters with abnormally low pH, caused by excessive sulfur content). Finally, spring outflows are often intermittent, creating yet another form of temperature heterogeneity. These intermittent flow communities vary greatly in their biotic composition, depending not only on the temperature gradient but also on the period and volume of flow, the chemical composition of the source water, and so on. The biodiversity found at temperatures less than 40°C is, however, low compared to the biodiversity of similar temperature zones in a relatively constant thermal gradient.

Thermal vent communities are formed by the emission of superheated water from fissures in the ocean floor that form along the line marking the meeting of tectonic plates. Because of the pressure of deep water, temperatures are possible that greatly exceed the 100°C temperatures of water boiling at sea level. Unfortunately, the great depths at which these vent communities form makes their study both difficult and extremely expensive. In general outline, the water is colonized by strains of high temperature heterotrophic bacteria utilizing the organic compounds dissolved in the superheated water. A strong thermal gradient is set up where the heated water meets the cold seawater and several groups of filter-feeding marine invertebrates have evolved to utilize these resources. Whether this evolution involves tolerance to high temperatures, however, is problematic, since samples and temperatures are hard to match up under such difficult sampling conditions. Furthermore, the thermal gradient is so steep that small distances may see a radical change in mean temperature. Numbers of individuals and biomass of the invertebrate consumers are large, because of the richness of the production by heterotrophic microorganisms, but number of species is low, although perhaps similar or even higher than that of the cold, dark, relatively sterile community away from the thermal vents at these depths. Although the microorganisms are clearly living at temperatures far higher than any found at the surface of the earth, the question of whether multicelled eucaryotes have evolved to survive at temperatures above 40 to 50°C is still an open question.

IV. TEMPERATURE LIMITS TO LIFE

Extreme environments were characterized earlier by the absence of species belonging to large systematic groups. In thermal ecosystems, as noted earlier, not only are these large systematic groups not found, but within groups, species numbers are low, compared with cool water communities. In general, in terms of temperature tolerance, there is a progression from more "primitive" to higher taxonomic groups, and from small to large individual size. Later I consider the arguments for the evolution of the various temperature tolerance ranges. What are the observed limits, based on sampling and observation of thermal ecosystems?

Four different groups of organisms need to be considered: (a) the heterotrophic procaryotic bacteria, (b) the procaryotic photoautotrophs, (c) the eucaryotic microorganisms, both heterotrophic and photoautotrophic, and (d) the eucaryotic metazoans. Much of the literature prior to 1978 on these limits has been summarized. The study of photosynthetically active thermally tolerant organisms really began with the pioneering work of W. A. Setchell. The main body of his work was never published, but it exists as a 215-page manuscript in the archives of the University of California at Berkeley. A brief summary of the temperature limits portion of this work was published in *Science*. Setchell reported the upper temperature for "algae" as 75 to 77°C, and for bacteria to be 89°C. A later investigator claimed, on the basis of superficial samples in Yellowstone, that organisms could not grow above 73°C, but this work relied on the uptake of radioactive phosphorous as the indicator of life; what was very likely being measured was the upper temperature for photosynthesis rather than the upper temperature for life. In other older records of maximum temperatures, unfortunately, some observers confused blue-green algae, procaryotic cyanobacteria, with green algae, eucaryotes. Within the procaryotes, there has also been confusion about cyanobacteria versus filamentous bacteria. Furthermore, it was not realized how steep the thermal gradient could be, thus putting a premium on temperature measurements at precisely the point where organisms are growing in the field. More recently, the discovery of ocean thermal vents (discussed earlier) has reopened some of the controversy regarding these limits.

Heterotrophic procaryotes have colonized habitats at all temperatures up to the boiling point of water (100°C) at sea level. The most thermophilic of these organisms have not been cultured, but they are easily sampled in the parent thermal communities using glass slides and simple photomicroscopy. On the basis of these findings, the prediction is that life is possible at any temperature at which there is liquid water. This prediction has now been verified by the preliminary explorations of the thermal vent communities on the ocean floor (discussed earlier) with microorganisms living at temperatures far in excess of 100°C. However, reports of organisms living in water to 300°C are still provoking argument. As noted earlier, the most extremely thermophilic procaryotes have not even been cultured. This may be in part due to insufficient knowledge about the nutritional requirements of these heterotrophs, but it might also be due to the difficulty of maintaining cultures near the boiling point of water. The problem is exacerbated in the case of thermal vent microorganisms growing at both extreme temperatures and pressures. Many of these species are also obligate anaerobes, for which even extremely low concentrations of oxygen are poisonous. Others are endosymbionts, which are also notoriously difficult to culture. Just getting the samples to the surface in a viable condition is a problem, and special culture techniques are required for the methanogens. In summary, thermophilic heterotrophic procaryotes are apparently not limited by temperature but by the presence of liquid water. At the surface this is approximately 100°C (depending on altitude), but at depth in the ocean life is found at substantially higher temperatures, limited by suitable nutrients, pH, and energy sources.

The temperature limit for procaryotic photosynthetic autotrophs (both cyanobacteria and photosynthetic bacteria) is substantially below 100°C on the surface. All of the observations and experimental evidence to date suggests a maximum limit of 73 to 74°C. Of this group the most thermophilic are the photosynthetic procaryotes, of which the most temperature tolerant is the single-celled species of cyanobacterium (*Synechococcus lividus*) and the filamentous photosynthetic bacterium (*Chloroflexus aurantiacus*). These species are found in nature at temperatures up to 74°C, although the optimum temperature is 63 to 67°C for the cyanobacteria and even lower, about 55°C, for the filamentous bacterium. In general, the filamentous forms of the photosynthetic procaryotes seem to have adapted to much lower optimal temperatures than the single-celled forms such as *Synechococcus* sp.

Eucaryotic microorganisms are found in nature at substantially lower temperatures than are procaryotes. Different reactions and responses to thermal environments are found in (a) fungi, (b) eucaryotic algae, (c) protozoa, and (d) metazoans, including invertebrates, vascular plants, and, vertebrates.

Fungi: In general, thermophilic and thermotolerant fungi are found at temperatures lower than 60°C and occur at this temperature only in acid thermal waters. In the very common and widespread "alkaline" hot spring communities, where the pH is initially somewhat acid because of dissolved carbon dioxide but rapidly rises when exposed to the air, free-living filamentous fungi are absent, even at temperatures below 40°C. There are reports of fungi parasitic on cyanobacteria that would probably be found at somewhat higher temperatures, but these have not been found in the "alkaline" thermal effluents of Yellowstone Park and are, in any case, poorly known (although there is a remarkable thermal range of the disease-causing fungus, *Dactylaria gallopava*).

Single-celled and filamentous green algae as well as diatoms become abundant at temperatures around 40°C and below. Earlier reports in the literature of diatoms growing at higher temperatures relied only on the recovery of the resistant siliceous frustules, without demonstrating that the cells were alive or growing. The maximum temperatures for which diatoms can be proved to be surviving and growing is 43 to 44°C. This particular diatom, *Achanthes exigua*, has an optimum temperature of 40°C. The single-celled and filamentous green algae have very similar temperature limits to those of the diatoms. Below 40°C in the "alkaline" thermal outflows of Yellowstone, the diversity of algal species rises (discussed in a later section). Although some mats of filamentous green algae may be found in these communities on top of mats of filamentous bacteria and cyanobacteria, where the underlying water is much hotter than 40°C, this is because the green algae is growing on top of the mat and is not exposed to the higher temperature water. This underscores the care needed when measuring the temperature of the environment of suspected thermally tolerant or thermophilic microorganisms. Strong gradients can occur in millimeters between where temperatures are measured and where the organisms are found. An interesting and important exception to the restriction of the eucaryotic algae to temperatures in the range of 40°C is that of the acidophilic single-celled asexual green alga, *Cyanidium caldarium*. This organism has been studied intensively in Yellowstone Park acid hot spring effluents. The upper temperature limit is 55 to 57°C, and the organism can be found down to 35°C. The high temperature adaptation of this acidophilic eucaryote has been explained as a simple matter of no other competition, but the story may be more complicated than that (see next section on the evolution of thermophily).

Protozoa have not been given the attention of other eucaryotes in terms of their occurrence in thermal environments. Determining the temperature limits of this group is difficult because of erroneous reports in the older literature reporting the growth and survival of protozoa at high temperatures. Protozoa will grow readily at temperatures up to 45°C. Other studies have reported somewhat higher temperatures (57 to 58°C seems to be the maximum), but the majority of protozoa in the natural thermal systems are found at lower temperatures. I have not found living protozoa in the "alkaline" thermal outflows of Yellowstone at temperatures greater than 40 to 43°C.

Invertebrates are ubiquitous in and around thermal outflows throughout the world. In surface thermal features many are winged and thus widely distributed. The same genera and species are found in widely separated thermal ecosystems. Some species, such as the water mites, have a juvenile stage parasitic on winged adult insects and are also widely distributed. Those invertebrates that are neither winged nor parasitic are more local in the distribution of species. Most invertebrate inhabitants of thermal ecosystems are motile, thus enabling them to choose, within limits, the space and temperature optimal to growth and survival. Those that are not motile, or only minimally so, such as many of the deep-sea thermal vent invertebrates and the eggs and juveniles of many winged insects are subject to mass mortality when major shifts in current and in the thermal gradient occur. To date, no invertebrates have been found surviving and growing at temperatures greater than the 50°C claimed for the ostracod, *Potamocypris*. The thermal limits of many of the most abundant and important invertebrate inhabitants of thermal ecosystems are in fact often quite low (well below 40°C). They exist by quickly colonizing temporary cool spots in the algal mats and growing into motile adults before the thermal environment becomes hot again. These are typical examples of thermal opportunists.

Vascular plants, being rooted in place, can colonize thermal waters at their maximum temperature only if the temperature is extremely stable. Thus there has been little study of their maximum tolerances. Species of *Juncus* grow in thermal streams in Yellowstone Park up to about 40°C and are often so abundant that they block stream flow and cause minor flooding of adjacent soil. Many species of vascular plant grow along the sides of the streams, one of the more conspicuous is the yellow monkey flower (*Mimulus guttatta*). It is difficult to say to what degree this plant is thermal tolerant because it does not grow directly in the hot water; its restriction in thermal basins to the sides of streams may simply reflect the availability of water, since the soil (sinter) beside thermal outflows is very dry. The species is found in wet swampy areas throughout Yellowstone Park and shows a wide range in size, being smallest in the hottest and least fertile environments.

Vertebrates, even more than vascular plants, are absent from thermal environments. One fish, the desert pupfish (*Cyprinodon*) is found in nonthermal (solar-heated) streams of Death Valley, California. Here it tolerates temperatures up to 43.5°C. But geothermal outflows are too hot, too small, too ephemeral, and vary too much in temperature and flow to support populations of even the small vertebrates. Reports of fish living in very hot water suffer from the same problem of most of the early observations of thermal maxima in organisms—namely, the means of temperature measurement were not sufficient to measure small gradients. Often the natural gradient of the water was such that the fish swam quickly from cool water into warm water to feed and then retreated to their cool refuges. For example, in Yellowstone Park, the many nutrients, particularly phosphorous, added to large, cold water rivers such as the Firehole, create exceptionally fertile and productive rivers. But these inflows also raise the temperature to the point that in summer they create a naturally thermally polluted river. In summer the lower Firehole river often increases in temperature far above the optimum for trout. For example, above the geyser basins in the summer of 1952, the average temperature was 10.2°C, whereas below the geyser outflows the temperature was 18.7°C, despite the influx of cold water from nonthermal springs and streams. In September of 1962 the difference was even more marked, 10.6°C above and 20.0°C below. That fish are able to not only survive but prosper in this situation is due to the many cold spring and streamside inflows entering the river, where the trout spend the majority of their time, venturing into the warmer currents of the river only to feed for short periods.

V. THE EVOLUTION OF THERMOPHILY

Thermophily in general must be thought of as the adaptation of organisms to survive and grow at temperatures (water as used in this chapter) significantly higher than the range of temperature in nonheated waters of the geographic region or climate. I have defined the range of thermal heating and the different types of ecosystem developing from the heated outflows. The types of organisms and their maximal and, where available, their optimum temperatures were discussed earlier. In general, the number of species and the diversity of the higher taxonomic classifications is lowest at the highest temperatures and increases with temperature decrease. The absolute upper temperature for life seems to be set only by the temperature at which liquid water is present. But if heterotrophic procaryotes can evolve to live and grow at these highest temperatures, why do not the other microorganisms and the metazoans also evolve to exploit higher temperatures? In one of the first (and the most detailed) discussions of the evolution of thermophily, the early arguments about this issue are reviewed and new hypotheses presented. Prior to the mid-1970s, the question of thermophily was regarded primarily as a physiological question, not an ecological one. This despite a wealth of evidence from other types of extreme environments that showed that adaptation to the extreme characteristic did not only have to be physiological, but that competition with other organisms was an important factor.

Several earlier authors regarded the increasing temperature tolerance of metazoa to eucaryotic algae to cyanobacteria (plus fungi) to heterotrophic bacteria as being due primarily to the increasing complexity of the cellular structure. In the fungal eucaryotes and for the eucaryotic green alga, *Cyanidium*, both of which reach 60°C as the upper maximum, the situation is complicated by the fact that they grow only in low pH hot systems. For the most part the arguments are simple statements that the cell structure has some fundamental property that cannot be overcome by further adaptation. This postulates not only that the fundamental limit exists, but that there are positive selective pressures for the further adaptation to high temperatures. To date there is little evidence for the idea that temperature tolerance is a simple function of cell complexity, nor has the locus of the effect been identified. At first it was thought that the deep sea invertebrates were extraordinarily tolerant of high temperature, and this might have stimulated research into the cellular question. However, recent evidence points to rather low temperature tolerances in these groups. Clearly, no animals have been found living in the "black smokers" where temperatures of 400°C or more are common. White smokers at about 50°C may have some fauna, but the majority of the vent fauna is distributed around the vents, in much cooler water at 40°C or less. Since these temperatures are far above the ambient temperatures of the deep sea, the fauna is, in the sense defined earlier, thermophilic, but will do little to settle the question of why eucaryotes have not evolved to tolerate temperatures endured by thermophilic procaryotes. The essential test of the cell complexity-membrane postulate was to separate the question into: (a) the evolutionary advantage to be achieved above a maximum limit of 62°C and (b) the evidence that the various eucaryotic genomes have the capacity to respond to the selective pressures if they exist.

Following the preceding argument, the reasoning is that adaptation of eucaryotes to high temperatures could be prevented by some inherent physicochemical limitation or by the lack of benefit in additional resources gained. Two tactics are possible: (a) to shift the range of temperature tolerance at the upper level (the maximum) independently of the lower range and (b) to shift the entire range of temperature tolerance upward. Thermal outflows vary considerably in width, depth, and flow volumes. But the nonlinear cooling process (more rapid at higher temperature differences between water and air/substrate) ensures that the total area of colonization space, and thus of resources (whether solar energy or dissolved nutrients), will become steadily smaller as one proceeds, for example, in 5°C increments, down the thermal gradient. Thus organisms that can adapt to higher temperatures by moving only the upper tolerance limit, keeping efficiencies at lower temperatures constant, will always have a positive selective pressure for doing so. If, however, the range must remain the same, shifting it upward will result in an overall loss of resource space. In the latter case, there will never be a positive advantage to be accrued in shifting the range unless other factors are involved, such as predation, competition, or the thermal range of a required food species. Therefore, any general explanation of the evolution of temperature tolerance by themophiles must consider not only a possible physicochemical limitation, but also whether resources are present and the relative amounts of resources at different temperatures. The first of these parameters, physicochemical limits due to cellular differences, has been the focus of most of the early literature. The presence of abundant resources for the various forms of eucaryotic organisms found in and around thermal ecosystems has been well established, to the point where it cannot be invoked as a sole or even a major block to the adaptation to higher temperatures. The invertebrate grazers of the algal mat are restricted to lower temperatures than their food organisms. In the case of photosynthetic eucaryotes, there is the same amount of solar energy per square meter at high temperatures and nutrients are at least as abundant as at low temperatures, if not more so. Thus the question devolves to physicochemical limits or amount of space and therefore of resources. To examine the effect of resource availability (read area) the two tactics discussed above are used. Only populations that are genetically homogeneous or whose genotypes are distributed by random mating fit the criteria. Asexual procaryotes do not fit because changes in genotypic abundance are the result of rates of growth and death. New genotypes arise by mutation and some exchange of genetic material, but the extent of the latter was not known in the 1970s and to my knowledge is not known today in thermophiles.

Because water begins to cool once it emerges from depth into the open air, surface thermal systems typically have the highest temperature water in spring-fed pools with relatively little current and little day-to-day or even annual variation in temperature. At temperatures above the (approximate) 75°C limit for photosynthetic autotrophs, these pools are dominated by monotypic or very low diversity heterotrophic procaryotes and the pools, although small in absolute area, are virtually infinite in size compared to the size of the bacteria colonizing them. These bacteria are often attached rods or sometimes filaments. Hyperthermophilic bacteria also occur in those thermal outflows where temperatures are 80°C or above. Because of the current, these forms must be attached rods or filaments. Furthermore, the colonies in the hot outflow channels decrease rapidly downstream. Whether this is because of the decrease in downstream temperature or the development of nutrient shadows is still an open question. Experiments in the field made by enriching these channels with nutrients suggest that the latter may be a more reasonable explanation than the former.

The dominant organism in the high temperature outflows, *Thermus aquaticus*, is an obligate aerobe that forms attached filamentous colonies at temperatures up to 80°C. Discovered and named in 1969, it now appears to be a complex of several different bacterial types. If the attached bacteria in very hot outflows do show changes in growth and viability with temperature, it might appear as a broadened range of temperature tolerance. However, this could also be explained as the successive downstream colonization of different genotypes, each adapted to a different optimal temperature. Alternatively, if competition for a scarce nutrient presents a major restriction on distribution, low temperature tolerance will be sacrificed to gain higher temperature tolerance. In this case successive adaptations to higher temperature result in a gain in resources providing the new genotype can maintain a sufficient level of efficiency to gain an advantage over its lower temperature-adapted competitor. Since I have suggested that the apparent temperature response could, in fact, be a response to nutrient depletion, the whole question is still an open one. These colonial forms can be cultured in the laboratory and also can be easily used in field experiments. Some efforts to solve the question of nutrient versus temperature would be worthwhile. In contrast, the extreme thermophilic bacteria that exist in the source pools at temperatures up to the boiling point are very refractory to culture. Thus studying their genetics and measuring their response to temperature clines has been impossible.

In the procaryotic photosynthetic microorganism, the evolution of temperature tolerance was operating in the same way, that is, the adaptation of successive genotypes to higher and higher temperatures, where there is a gain in resources, should be observed. Instead, we observe the complete failure of these organisms to colonize water at temperatures greater than 73 to 75°C (see discussion on limits presented earlier). The two major resources of the photosynthetic cyanobacteria and the flexibacteria are sunlight and a carbon source (phosphate is abundant in these systems and the organisms fix nitrogen). Cyanobacteria in particular respond positively to increases in free carbon dioxide. Sunlight is intense at the Yellowstone altitude in all shallow-water systems, obviously independent of temperature. Sunlight only becomes limited deep in thick algal mats, thus expanding the thermal range upward by the development of higher-temperature clones will neither increase or decrease the availability of this factor to thin new mats of cyanobacteria. In the "alkaline" thermal flows however, the water typically emerges at a pH in the 6 to 6.5 range due entirely to large concentrations of dissolved carbon dioxide. Therefore, any organism capable of forming genetically homogeneous clones should be able to adapt to higher and higher temperatures by gaining resources (carbon dioxide) at hotter temperatures. Such clones are known for both of the most common genera of cyanobacteria in thermal springs, *Mastigocladus* and *Synechococcus*.

The best evidence for an inherent physicochemical limit on the temperature of photosynthesis would be if the previous conditions were met with respect to resources. But adaptation proved unable to extend the range of temperature tolerance without a significant decrease in the ability or efficiency to use the increased abundance of the resource at higher and higher temperatures. This is exactly the case for *Synechococcus*. When growth (doubling time) is measured as a function of temperature, it is five times longer at 70°C than at 40°C. Thus adaptation to higher temperatures results in a rapidly decreasing ability to process a limiting nutrient, even though it (presumably carbon dioxide) is more concentrated at the higher temperatures. Thus the failure to evolve tolerances above 73 to 75°C must be due to a temperature dependent reduction in growth rate or a physicochemical failure in the algal cell.

Eucaryotic microorganisms, including single-celled and filamentous forms plus diatoms and free-living fungi, are seldom found at temperatures above the range of 40 to 45°C, with the exception of some fungi and the acidophilic genus *Cyanidium*. Collections made from six different hot springs in Montana showed a clear and significant separation between the eucaryotic algae and the cyanobacteria; the former were never collected at temperatures greater than 50°C and the latter showed more species per sample (2 to 5 at 50 to 58°C, 10 at 35°C) than the former. But the diversity of the eucaryotic algae was increasing much faster than the cyanobacteria with decreasing temperature. This separation was present in the thermal springs of both Yellowstone and Mount Rainier National Parks. The eucaryotic alga, *Cyanidium*, which exists at much higher temperatures than the other eucaryotic algae, may be subject to quite different selection pressures since it also occupies a variety of adjacent habitats (wet soil) where the temperatures fluctuate greatly, as do the temperatures in the outflows, which are very cool during snowmelt in the spring and the alga disappears, only to reappear in the water when the temperature rises. The argument is made that the evolution of thermophily in this group must take into account these differences plus the fact that hot acid flows have no procaryotic photoautotrophs that could offer competition to *Cyanidium*. The possibility of genetic mixing due to sexual selection in the eucaryotic algae may, together with the increasing area for colonization as the thermal gradient is descended, render the selective optimum temperature lower in eucaryotes than in sets of competing procaryotic clones. This idea needs testing along with the notion of a major physicochemical limit on the eucarytotic cell. The growth of *Cyanidium* at higher temperatures certainly casts some doubt on the latter as a general explanation for the failure of most eucaryotic algae to adapt to temperatures higher than 50°C.

Metazoans typically have the lowest maximum thermal tolerances of any group of organisms. Here again, the arguments prevailing before the mid-1970s detailing the failure to adapt to higher temperatures invoked some vaguely defined physicochemical limit. In the case of the metazoa, complexity of organization is added to the supposedly inherent limits of the eucaryotic cell. But, as pointed out for other groups, this is only one of the three lines of evidence that must be considered when pondering the origins of thermophily. The other two are the availability of resources, both space and nutrients, and the interactions of resource abundance and evolutionary mechanisms that act to determine fitness. In other words, whatever the physicochemical limit to the evolution of thermal tolerance, simply measuring maximum tolerance rates in the field does not provide evidence, necessarily, that the physicochemical limit has been reached. Clearly, the larger aquatic metazoans could not sustain viable population densities in the small thermal systems, so there would be no selective pressure for them to colonize even moderate to low temperature thermal systems. Thus no fish are found in geothermally heated waters. The few cases of the evolution of tolerance to elevated temperatures in fish involve small individual size and solar-heated habitats. Here the maximum temperatures tolerated somewhat exceed 43°C.

The metazoans characteristic of hot springs are the arachnids (water mites), insects (mostly flies), small mollusks (small pulmonate snails), odonates (dragon flies), and, occasionally, ostracods. These groups are neither characteristic of the cold water streams into which the thermal effluents drain, nor of running water systems in general. Rather, they are related to (sometimes the same species as) the invertebrates found in the shallow muddy margins of lakes and ponds, where ambient daily maximum temperatures, due to solar heating, are of the same order as those in which the thermal inhabitants are found. If these low to moderate temperatures are not set by a physicochemical threshold, then the problem is to explain why, in thermal habitats, the species have not evolved more thermal tolerance.

The second factor affecting distribution (mentioned earlier), competition for scarce resources, seems not to be a factor. In every case studied, the grazers on the algae of thermal ecosystems are less tolerant of high temperature than the cyanobacteria on which they feed. Examples are the ostracods in Hunter Hot Springs, Oregon, and the ephydrid flies of the "alkaline" thermal flows in Yellowstone Park. This leaves the interaction of resources and selection determining fitness. Under random sexual mating (with respect to temperature), the direction of the evolution of temperature tolerance will be determined by the relative numbers of surviving offspring from the upper half of the temperature range relative to those from the lower half. But because of the nonlinear cooling curve of water, the size of the upper temperature part of the range will always be smaller than that of the lower half. Thus under a tactic of shifting range, thermophily cannot evolve unless there is nonrandom mating or fecundity or survival is lower at low temperatures. The randomness of mating with respect to temperature differs in the various groups of hot springs invertebrates, depending on life history and behavior. In the flies and dragonflies, mating occurs after emergence and the animals are highly motile, so randomness with respect to temperature of development is assured. The water mites are also highly dispersed as parasitic larvae, but they do spend their adult lives within a relative narrow temperature range and mate there. The smaller mite, *Partnuniella*, tends to spend most of its adult life in cooler patches of the cyanobacteria/flexibacteria mat where it feeds on the eggs of the ephydrid flies and, as expected, has a rather low temperature tolerance. The large predaceous mite, *Thermacarus*, lives directly in the hot water and has an upper temperature tolerance 5 to 10°C higher than the smaller species of mite and just lower than the most temperature tolerant metazoan, the ostracod. This group has no widespread dispersal mechanism, thus nonrandom mating is predicted. They also seem to be limited by food. Not surprisingly, these organisms have evolved the highest tolerance to high temperature of any metazoan.

In summary, current evidence supports the following statements concerning the evolution of thermophily: (a) the heterotrophic procaryotic bacteria seem not to have any physicochemical limit, having evolved to inhabit geothermal systems of any temperature as long as liquid water is present; (b) the photosynthetic cyanobacteria seem to have evolved to the point

(73 to 75°C tolerance) where a physicochemical limit has been reached; and (c) eucaryotes vary greatly in temperature tolerance, depending on the group, but none has evolved tolerances equal to the maxima exhibited by the procaryotes. In most cases, the observed limits can be explained on the basis of simple selection/fitness processes, without invoking unknown physicochemical mechanisms.

VI. THE BIODIVERSITY OF THERMAL ECOSYSTEMS

Species in, using, or around (living from) geothermal ecosystems throughout the world have been the subject of many studies. The earliest intensive work was on the fauna of the hot springs of Iceland. The algae of several hot springs in western India were cataloged in the 1960s. The algae and fauna of warm springs in New Zealand have also been studied, as has the fauna of thermal springs in the Dutch East Indies. The thermal effluents of Yellowstone National Park in the United States have been the subject of both the largest number and most intensive studies of thermophiles of any area in the world. The original studies concentrated on the fauna but also included some observations of the algae (cyanobacteria). Beginning in 1964, detailed studies of the microorganisms living in the hot spring waters of Yellowstone Park were begun. These became both the most extensive survey and intensive experimental protocol yet done on thermal microbiology. At first the focus was on high temperature systems where only one or a few species were found in any given thermal ecosystem. In 1967, I and my colleagues and students began studying the lower temperature thermal ecosystems with higher diversities of microorganisms and a richer food chain involving species of animal acting as scavengers, predators, and parasites. Additional unpublished information has been obtained from 1983 to the present on the thermal ecosystems developing at temperatures below 45°C. Thus the ecology of the Yellowstone ecosystems can begin to be understood from the standpoints of species diversity and food-chain relationships. In this final section I review what is known about the diversity of these systems, in their various manifestations of temperature and pH, and compare the results with the thermal biodiversity of the deep sea thermal vents, insofar as the latter are currently understood.

High temperature is the most difficult of extreme environmental conditions to which organisms can attempt to adapt. There are many ways to adapt to exist in extreme cold or aridity and resume growth and activity during short seasonal changes in solar warming and precipitation. High salinity can be avoided or salt excreted so extremes can still support a reasonably high diversity. Indeed, many saline ecosystems may appear simple from the standpoint of vascular plant species and the food chains they support, but be relatively high in the diversity of algae and bacteria in the surface sediments. Thus, in every thermal system investigated to date, the biodiversity decreases rapidly with increases in temperature. At their most diverse, thermal system species are measured in the hundreds; their least diverse manifestations are monospecific.

Temperatures near the boiling point of water in Yellowstone thermal systems and elsewhere (85°C–100°C) are most common in source pools fed by deep subterranean springs of hot water. These pools vary widely in chemical characteristics, depending on the mineral content of the rocks through which the superheated water passes on its way to the surface. All such pools have a large volume of heated water with a relatively small inflow/outflow. Currents are small; almost all mixing is by convection or mild bubbling of various gases, often carbon dioxide. Such systems are natural analogues of chemostats, and they are eventually colonized by adapted heterotrophic bacteria (photosynthesis being impossible at these high temperatures). Not surprisingly, competition generally ensures that they are monospecific, or, if chemoautotrophy is possible, the diversity might rise by one or two additional species. Although thermal source pools may be thought small relative to surrounding mountains, forests, and so on, from the viewpoint of microorganisms, these are virtually constant environmental systems capable of supporting immense numbers of individuals. Current ecological theory supports the prediction of dominance by one or a few species in these circumstances. In contrast, the emergence of superheated water from the thermal vents in the deep sea cannot reproduce these constant conditions. Pools of hot water do not exist, but instead a turbulent mixture of water is emitted at temperatures apparently as high as 350°C or more, directly into cold ocean water near 2°C. Thus environmental variation in temperature is undoubtedly high and the higher diversity of microorganisms would be predicted. It is impossible, under such variability in temperature, for one species to monopolize the resources of the system. Thus the students of thermal vent microbiology are finding a number of thermophilic species new to science (although no novel genera have been found). The latter may to due to the culture techniques used, which have not been designed for organisms adapted to high pressures as well as high temperatures.

Food chains in the deep sea vent communities are simple and short, because in the absence of light no photoautotrophy is possible. Thus the vents are surrounded by a diverse array of marine invertebrates that ingest the cells and products from the thermal vents. Almost all of these animals are novel species. As many as 236 species of animals have been collected around thermal vents, of which 223 are new to science. This diversity of animal species is far higher than that of any single category of surface thermal ecosystems, although the data represent many different thermal vent locations. If the worldwide diversity of animal species for all known surface thermal features is used, the animal diversity per system is similar in the two types of system. Thus the thermal vent communities of the deep sea, although of great interest from the standpoint of physiological adaptations, mass mortality, and colonization strategy, appear to offer little that is new to ecological theory.

At temperatures below 75°C and above about 40°C to 45°C, thermal outflows form small streams of hot water. These are colonized throughout the world by unicellular photoautotrophic cyanobacteria and filamentous bacteria. These biotic communities are characteristic of all surface thermal features throughout the world, but most early studies were simply endeavoring to determine maximum temperatures. The first rigorous studies of these systems were initiated in Yellowstone Park. In each ecosystem, there are generally two dominant species or species groups at the upper end of this temperature range (a base mat of the filamentous photosynthetic bacterium, *Chloroflexus,* and an overlying film of the unicellular cyanbacterium, *Synechococcus*). As the temperature decreases, the overlying mat becomes dominated by one or more genera of filamentous cyanobacteria. In relatively constant flows, temperature-adapted strains of *Mastigocladus* are the common dominants; in springs with variable ouflows, diversity is often higher due to the presence of additional genera of filamentous cyanobacteria, commonly *Calothrix* or *Phormidium*. When these filamentous forms are able to grow, the outflow begins to go through a definite cycle, because the filamentous mat is not limited to a thin film as are the unicellular forms. As the mat grows, it thickens unevenly; water flows around these thickened areas, producing cool patches. These patches are populated, almost immediately, by one or more species of brine fly (Ephydridae). Upon hatching, the fly larvae begin to eat and destroy the filamentous integrity of the mat. The pupae metamorphose to adults within a few days. Eventually hot water reenters the cool patch, remaining larvae and pupae are killed, and the cycle starts over in another part of the mat. These flies and their predators are not thermophilic, and in fact are rather thermally intolerant, relative to organisms living in other thermal ecosystems—compost piles, for example. They are an example of what I have termed thermal opportunists, organisms that have adapted their life cycle to take advantage of opportunities provided within the thermal ecosystem. The diversity of the alkaline filamentous mat-fly association is not high for most expressions of this community of 20 to 30 species. Yet a diverse set of food chains and ecological processes is found, photosynthesis, grazing, decomposition, and predation, for example.

When a strongly channeled outflow cools down to 40°C or below, the diversity goes up rapidly. Metazoan animals can adapt to live directly in the warm water, and eucaryotic algae such as green algae and diatoms invade. Although cyanobacteria are still abundant, they are heavily grazed, and do not form a thick mart. In different expressions of this community are found crustacea, midges (Diptera), and water mites (Hydrachnellae), dominated by a top predator, the larva of the dragonfly, *Erythemis collocatta*. The net effect on diversity is to increase the species numbers to 50 to 60, exclusive of the nonphotosynthetic microorganisms. The latter have not been studied in detail at these temperatures, thus a total diversity of 100 or more would not be unexpected.

VII. SUMMARY

Thermal waters, as extreme ecosystems, exhibit the low diversity expected under such conditions. But high temperature is a particularly difficult factor for adaptive evolution to overcome. Thus, thermal ecosystems appear to have the lowest diversities of any of the extreme environments. Nevertheless, there is considerable variation in diversity, depending on both the temperature of emerging water, the rate of cooling, and the stability of the emerging plume. Generally, there is an increase in diversity as the temperature declines. This can be modified by short-term fluctuations in temperature caused by currents set up in the emerging plume of hot water (in deep sea thermal vents) or by intermittent flows from the source creating cooler periods downstream. Both of these effects increase diversity. When animals are able to enter the system, a rich ecological food chain (in terms of numbers of ecological processes) often develops.

Acknowledgments

I am indebted for the information in this article to dozens of previous studies of the ecology of thermal organisms. My own work in Yellowstone National Park and elsewhere has benefited from permission from the park administration to do research. I have received financial support from the National Science Foundation through a series of grants, the latest of which, DEB-821-0142, was used to investigate the community forming at temperatures below 40 to 45°C. I am also indebted to Drs. A. Chalmers and R. Hodson, who provided useful comments on the manuscript. Support for preparation of the manuscript was received as discretionary research funds from the University of Georgia.

See Also the Following Articles

PSYCHROPHILES • THERMOPHILES, ORIGIN OF • VENTS

Bibliography

Brock, T. D. (1967a). Life at high temperatures. *Science* 158, 1012–1019.

Brock, T. D. (1967b). Microorganisms adapted to high temperatures. *Nature* 214, 882–885.

Brock, T. D. (1969). Microbial growth under extreme conditions. *Symp. Soc. Gen. Microbiol.* 19, 15–41.

Brock, T. D. (1978). *Thermophilic Microorganisms and Life at High Temperatures.* Springer-Verlag, New York.

Brues, C. T. (1924). Observations on animal life in the thermal waters of Yellowstone Park, with a consideration of the thermal environment. *Proc. Amer. Acad. Arts and Sci.* 59, 371–437.

Brues, C. T. (1927). Animal life in hot springs. *Quart. Rev. Biol.* 2, 181–203.

Brues, C. T. (1932). Further studies on the fauna of North American hot springs. *Proc. Amer. Acad. Arts and Sci.* 67, 185–303.

Castenholz, R. W. (1973). The ecology of blue-green algae in hot springs. In *The Biology of Blue-Green Algae* (N. G. Carr and B. H. Whitten, Eds.), pp. 379–414. Blackwell, Oxford.

Childress, J. J., and Fisher, C. R. (1992). The biology of hydrothermal vent animals: Physiology, biochemistry and autotrophic symbioses. *Oceanogr. Mar. Biol. Annu. Rev.* 30, 337–441.

Collins, N. C., Mitchell, R., and Wiegert, R. G. (1976). Functional analysis of a thermal spring ecosystem, with an evaluation of the role of consumers. *Ecology* 57, 1221–1232.

Kullberg, R. (1971). Algal distribution in six thermal spring effluents. *Trans. Am. Micr. Soc.* 90, 412–434.

Lutz, R. A., and Kennish, M. J. (1993). Ecology of deep-sea hydrothermal vent communities: A review. *Reviews of Geophysics* 31, 211–242.

Mitchell, R. 1974. The evolution of thermophily in hot springs. *Quart. Rev. Biol.* 49, 229–242.

Prieur, D. (1971). Microbiology of deep-sea hydrothermal vents. *Marine Biotechnology, TIBTECH* 15, 242–244.

Segerer, A. H., Burggraf, S., Fiala, G., Huber, G., Huber, R., Pley, U., and Stetter, K. O. 1993. Life in hot springs and thermal vents. *Origins of Life and Evolution of the Biosphere* 23, 77–90.

Tunnicliffe, V. The biology of hydrothermal vents: Ecology and evolution. *Oceanogr. Mar. Biol. Annu. Rev.* 29, 319–407.

Wiegert, R. G., and Fraleigh, P. C. (1972). Ecology of Yellowstone effluent systems: Net primary production and species diversity of a successional blue-green algal mat. *Limnol. Oceanogr.* 17, 215–228.

Wiegert, R. G., and Mitchell, R. (1973). Ecology of Yellowstone thermal effluent systems: Intersects of blue-green algae, grazing flies (*Paracoenia*, Ephydridae) and water mites (*Partnuniella*, Hydrachnellae). *Hydrobiologia* 41, 251–271.

HISTORICAL AWARENESS OF BIODIVERSITY

David Takacs
California State University, Monterey Bay

THIS ARTICLE DISCUSSES how we have come to know and understand the subject of this encyclopedia. The term "biodiversity" was coined in 1986 by biologists who wished to express a complicated, scientific understanding of the natural world, and who wished to inspire a rapid, widespread effort to conserve the natural world. This article traces how biologists understand "biodiversity" and how they have attempted to raise awareness on its behalf.

I. BIOLOGISTS AND BIODIVERSITY BEFORE 1986

Since ancient times, scholars have simultaneously revered the natural world, attempted to discover in that world a natural order or impose rational order on that world, and sought to understand the place of humans in the cosmos based on what they read in the natural world.

Worster (1987), Bowler (1993), and other historians offer extensive treatments of pre-20th century attempts by biologists and their intellectual predecessors to understand, categorize, and philosophize the natural world. In the mid 18th century, for example, in order to fathom God's wisdom, Carolus Linnaeus classified the riot of life into a functional taxonomy, the *Systema Naturae*. In "The Oeconomy of Nature," (1749) Linnaeus offered a protoecological treatise on how these life forms fit together in mutual, stable interdependence, with humans at a central nexus in the web. For Linnaeus and his contemporaries, God meant humans to use those species that we found valuable; the complexity and richness of the natural world meant that our use of nature could not upset what God had balanced so exquisitely.

A century later, Charles Darwin built on Linnaeus's and others' ecological and taxonomic groundwork and painstakingly laid out a theory that provided an historical, deterministic explanation for the relatedness of all God's creatures—even if his work made God seem a bit less relevant and humans a bit less central to the machinery of nature. Darwin saw all forms of life as human kin and believed a key to civilization's maturity was the ability to empathize with and respect our extended family. For both Linnaeus and Darwin, the diversity of earth's life forms was a phenomenon to be revered, contemplated, and explained. Neither envisioned the object of their study as a commodity that needed protection from human advances, although a contemporary of Darwin's, George Perkins Marsh

(1864), did begin to grasp this in his prescient *Man and Nature*. Marsh warned that human activity was destroying natural functions, and that this destruction threatened to undercut the survival of our own species.

In the 20th century, some biologists looking for meaning in the diversity of life added to the trio of reverence, imposition of order, and search for deeper meaning about humanity. Biologists came to understand and document that the grand panoply of life forms was threatened as the fabric of ecological connections was rent. This awareness led biologists to make leaders and laypersons aware, and they searched for ways to use their carefully tended scientific authority to make nonscientists care about and conserve the objects of their reverence. Such work carries perils: advocacy threatens to undermine the perception of value neutrality and objectivity that leads laypersons to listen to scientists in the first place.

In *A Sand County Almanac II* (1970) wildlife ecologist Aldo Leopold proposed a "land ethic" that we should follow if we are to revere nature's diverse organisms and protect them to play the precise roles in nature's scheme that ecological scientists were just beginning to elucidate: "If the biota, in the course of aeons, has built something we like but do not understand, then who but a fool would discard seemingly useless parts? To keep every cog and wheel is the first precaution of intelligent tinkering" (p. 190). Leopold asserted that all species are intertwined in complex interrelationships, and the diversity of organisms and their interrelationships are crucial for a stable, functioning planet. From the "is" of ecology, Leopold derived the "ought" that forms the basis of his land ethic: "A thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise" (p. 262).

British ecologist Charles Elton also viewed "ecological variety" as a threatened commodity. In *The Ecology of Invasions by Animals and Plants* (1958), he foretold the dangers humans face as anthropogenic introductions of nonendemic species sweep the globe and threaten the flora and fauna that had previously been left to evolve in splendid isolation: "It is not just nuclear bombs and wars that threaten us, though these rank very high on the list at the moment: there are other sorts of explosions, and this book is about ecological explosions" (p. 15). Like Elton, Rachel Carson (1962) warned that people could no longer take this variety for granted as benign backdrop to human affairs. Rather, ecological science taught that species diversity is essential for ecological and human health and that uncontrolled pesticide use threatened this diversity. Biologists like Leopold, Elton, and Carson still revered "natural variety," still sought to reveal the secrets of nature's order. But unlike pre-20th century natural philosophers, they desired that their readers come to view species diversity as a quantifiable, measurable entity that was inextricably bound up with their own well-being. Species diversity was now a threatened commodity, and humans who caused this threat were simultaneously threatened as that commodity diminished.

II. CREATION OF THE TERM "BIODIVERSITY"

Throughout the 1960s and 1970s, biologists fomented public alarm over the deteriorating environment. Prominent biologists (e.g., Ehrenfeld, 1981; Ehrlich and Ehrlich, 1981; Myers, 1979) helped raise awareness that diversity was a threatened commodity, and U.S. legislatures responded with laws designed to assuage the threat. Most notably, the 1973 Endangered Species Act recognized and sought to protect the paramount value of species diversity (Kohm, 1991).

Still, biologists were frustrated that efforts to protect diversity were not keeping pace with the furious rate of destruction. Walter G. Rosen, a biologist and senior program officer at the National Research Council (which advises the National Academy of Sciences [NAS]), brought together prominent scientists from the NAS with the clout of the Smithsonian Institution to host the National Forum on BioDiversity in 1986. At the National Forum, biologists and others concerned about imperiled diversity staged a consciousness-raising event that sought, and received, widespread attention from the public. Rosen coined the neologism "biodiversity" for the event as a convenient shorthand, a buzzword that would at once encapsulate biologists' understanding of a chaotic, diminishing natural world, and would raise public awareness about threats to the natural world (Takacs, 1996). In *BioDiversity* (Wilson, 1988), the collection of essays that chronicled the National Forum, Paul Ehrlich, Daniel Janzen, Tom Cade, Lester Brown, Michael Soulé, and other scientists declared the need to rouse public attention on behalf of biodiversity and exhorted their colleagues to adopt that mission.

As biologists promote the term and the complex worldview it represents, "biodiversity" has become a widespread conservation buzzword. Biologists write about it in scientific and popular presses, both exploring its complexity and advocating its protection. Environmental groups focus on it in fundraising efforts; conferences convened in its name occur regularly. Laypersons have joined biologists in attempting to shape the planet's physical, political, and normative environments to make more room for biodiversity.

III. PROBLEMS WITH OTHER FOCI OF CONSERVATION EFFORTS

Why has "biodiversity" gained prominence as a conservation buzzword, and why have biologists speaking on its behalf had some success in shaping public opinion about threats to the natural world?

Various scholars (Cronon, 1995; Evernden, 1992; Williams, 1980) note that nature is so all-encompassing that what one attributes to it may say more about the speaker than it does about the natural world. Previous attempts to preserve "nature" or "wilderness" are too vaguely defined or deemed elitist in some quarters. For example, Guha (1989) sees traditional advocacy for nature or wilderness preservation as setting the aesthetic desires of the rich against the needs of the poor who need land to survive. Guha's critique has, to some extent, been incorporated into the biodiversity preservation discourse. For example, as people came to appreciate the value of cultural diversity, they might also see diversity in all its forms as a normative good, particularly when efforts to preserve both may be mutually reinforcing (Nabhan, 1997). A document prepared for the 1992 United Nations Conference on Environment and Development notes that "cultural diversity is closely linked to biodiversity. Humanity's collective knowledge of biodiversity and its use and management rests in cultural diversity; conversely, conserving biodiversity often helps strengthen cultural integrity and values" (Reid *et al*, 1992, p. 23).

Prior to the advent of biodiversity, the most effective conservation efforts in the U.S. focused on endangered species. Biologists' foci changed because biodiversity represents a more sophisticated ecological worldview, and a more sophisticated view of what biologists want preserved and how they want it preserved. Biologists promoting biodiversity conservation also seek to circumvent certain problems arising from efforts to preserve endangered species.

Even though human activities accelerate the rate of species extinction, some opponents of conservation argue that species extinction is a natural process that we should let proceed. Biologists have difficulties defining with precision what constitutes a species (or a subspecies, like the Northern Spotted Owl) in the first place. We are not aware of the economic or ecological benefits some individual species confer, so it is difficult to argue for their conservation in some circles. Many of the insects, bacteria, plants, and other members of what E. O. Wilson (1987) calls "the little things that run the world" remain unidentified and unloved; invertebrates comprise only a tiny percentage of organisms protected by the U.S. Endangered Species Act. Land set aside to protect individual endangered species may prove insufficient if global warming induces species migration (Peters and Lovejoy, 1992). When we focus on species diversity, we sometimes ignore genic, population, community, or ecosystem diversity. Endangered species conservation proves nearly impossible in poorly explored areas, in oceans, and in nations without species checklists. Focus on species that are endangered can be a last minute emergency effort to save species that may no longer be playing functional ecological roles, and this strategy may be inferior to proactive efforts to preserve healthy populations in healthy ecosystems. Finally, some view the U.S. Endangered Species Act as a mixed blessing: its unyielding allegiance to species on the verge of extinction is also a political lighting rod that leaves little room for compromise and has set many citizens against conservation efforts (Mann and Plummer, 1995).

IV. WHAT IS "BIODIVERSITY"?

Elsewhere in this encyclopedia, you may read about definitions of biodiversity. In research I conducted (Takacs, 1996), I asked prominent biologists to define the term. Little, if anything, in the natural world is excluded from these definitions of biodiversity. Responses included that biodiversity is "the complete array of organisms, biologically mediated processes, and organically derived structures out there on the globe" (Jerry Franklin); "the whole package of genes, populations, species, and the cluster of interactions that they manifest" (Daniel Janzen); "the total number of genetic lineages on earth" (Thomas Eisner); "shorthand for all the richness of life" (Reed Noss); or "the sum total of plants, animals, fungi, and microorganisms in the world including their genetic diversity and the way in which they fit together into communities and ecosystems" (Peter Raven). According to Terry Erwin (1991, p. 3), it is "the sum of earth species including all their interactions and variations within their biotic and abiotic environment in both space and time."

So biodiversity represents a rich, complicated vision of life and a corresponding rich, complicated vision of what biologists want to see preserved. Although few biologists wish to see any species slip into extinction, what they really want—and what biodiversity really represents—is preservation of ecosystems, of vast stretches of land where the evolutionary process may continue relatively unfettered.

But ecosystems are hard to delineate, and even mentioning "evolution" lights a powder keg in some quarters. Kellert (1986, 1996) notes that the public is most likely to rally around "cognitively meaningful" organisms. Habitat has no fur, and ecosystems lack big, expressive eyes that rouse public affection and therefore attention to conservation causes. With biodiversity, biologists can focus attention on the charismatic megavertebrates the layperson finds endearing. These organisms are emotionally and ecologically appealing to the conservation minded: they often are at trophic levels that require large territories to survive. When we preserve the large areas of land they require, we also fortuitously protect a myriad of species and enable continued functioning of ecological and evolutionary processes that biologists value.

Conservation efforts on behalf of biodiversity enable multiple images and multiple strategies to protect and conserve the natural world; many of those strategies place more power in the hands of biologists to realize their conservation values. If definitions of biodiversity seem complex and all encompassing, that is part of why it has been successful as a conservation tool. At once it represents the complexity of biologists' worldview, the whole span of what biologists wish to conserve, and in it each of us can see that part of nature we cherish.

V. BIOLOGISTS AND THE PROMOTION OF BIODIVERSITY

When biologists speak about biodiversity, they simultaneously refer to small parts of the ecological world, the interrelatedness between those parts, the ecological processes that sustain those parts, and the evolutionary processes that gave rise to those parts. While appearing as a purely scientific, objective entity, biodiversity also encompasses political arguments on behalf of conservation and symbolizes much that biologists do not know about the natural world.

Biodiversity represents ecological complexity that, the more they study, the more biologists realize they do not understand. Biologists do not know, within an order of magnitude, how many species live on earth, so they cannot specify how many we are losing and cannot say what the loss of these species represents to ecosystems or to humanity's future prospects (Stork, 1997; Tilman, 1999; quotes in Takacs, 1996). Biologists believe that if people are alarmed by such ignorance, they will simultaneously take precautions about dismantling an ecological world that supports us in myriad unknown ways, will support biologists' research efforts to understand that world, and will listen to biologists' policy prescriptions on how and why we ought to save that world. For example, the question that E. O. Wilson (1992) is most frequently asked about the diversity of life is: "if enough species are extinguished, will the ecosystem collapse, and will the extinction of most other species follow soon afterward? The only answer anyone can give is: possibly. By the time we find out, however, it might be too late. One planet, one experiment" (p. 182). In light of this uncertainty, Wilson urges "prudence. We should judge every scrap of biodiversity as priceless while we learn to use it and come to understand what it means to humanity" (p. 351). In Wilson's view, biologists hold the key to that understanding.

Many who have come to call themselves conservation biologists cite this uncertainty as they attempt to have the public become aware of biodiversity and become concerned about its diminution. They warn that what we don't know about biodiversity will hurt us, and urge us that if we were to come to know it, our lives might have greater meaning. They seek a new ethic where biodiversity would be cherished. Aldo Leopold (1949) sought this, and his "land ethic" laid the groundwork for a new way of valuing diversity. Leopold knew that "no important change in ethics was ever accomplished without an internal change in our intellectual emphasis, loyalties, affections, and convictions. The proof that conservation has not yet touched these foundations of conduct lies in the fact that philosophy and religion have not yet heard of it." Today, a number of prominent biologists seek to have our intellectual emphases, loyalties, affections, and convictions match theirs. Some biologists see this as a crucial moment in history when they must help raise awareness of biodiversity; they see it as their responsibility to speak out about what they study and love so that subsequent human and natural history will reflect their values. For example, Janzen (1986) exhorts his fellow biologists: "Set aside your random research and devote your life to activities that will bring the world to understand that tropical nature is an integral part of human life."

Janzen is part of the "mission-oriented" discipline of conservation biology, which was founded not only to study biodiversity, but also to embrace its many values and to promote findings that will help conserve the objects of study. Built into the discipline's foundations are normative principles or, as discipline founder Michael Soulé (1985, p. 730) says, "value statements that make up the basis of appropriate attitudes towards other forms of life—an ecosophy." The pages of the discipline's flagship journal, *Conservation Biology*, report not just the latest research findings; they are also filled with thoughtful articles on conservation education, policy implications of research, and spirited debates on what types of activism are required or appropriate for practitioners. While some biologists feel they must advocate on behalf of the natural world they study and love, others fear that such advocacy threatens the scientific enterprise: if scientists are no longer perceived as objective and value neutral, why would societies fund their work and listen to their results?

Many scientists ignore the risks of advocacy and proselytize on behalf of biodiversity. Some well-known biologists (e.g., E. O. Wilson, Peter Raven, Thomas Eisner, Thomas Lovejoy) have spoken before Congress on the value of biodiversity. Others testify in courts as expert witnesses on behalf of biodiversity. Paul Ehrlich and others appear on radio and television to raise awareness about biodiversity. Many biologists write for the popular press or speak to general audiences on the values of biodiversity. They talk to garden clubs, and preach in church pulpits. Thomas Lovejoy takes senators, movie stars, and other influential people to the Brazilian Amazon so that they may experience biodiversity "in a way words can't touch. . . . And it has never failed to be a truly touching experience for them" (Takacs, 1996, p. 153). By helping laypersons become aware of biodiversity's beauty and fragility—by tireless efforts to, in the words of the "Father of Biodiversity" E. O. Wilson, "educate, educate, educate" (quoted in Anonymous, 1994)—biologists hope for widespread transformation of values. That transformation of values might lead to what Paul Ehrlich (1985) calls a "quasi-religious transformation" of feelings toward the wonders of the natural world, which would presumably be translated into action to conserve those wonders.

Many prominent North American biologists work in farflung corners of the earth to raise awareness of biodiversity, particularly in tropical nations that hold the greatest concentrations of diversity, and which face the greatest imminent threats to that bounty. In Costa Rica, a small country with perhaps 5% of the world's species diversity, biologists have spurred a national effort not only to preserve biodiversity in protected areas, but also to catalog biodiversity and to search for the economic wealth contained within. At Costa Rica's Instituto Nacional de Biodiversidad, specimens of plants, insects, and other taxa are sorted and sent to multinational pharmaceutical companies for research. Profits from successful drugs deriving from Costa Rica's biodiversity would be funneled back to the areas where the biodiversity flourishes. In effect, this is an experiment in which biodiversity would pay its own way, with biologists driving the process.

Specimens are collected by a large group of rural-dwelling parataxonomists. As they tromp around the nation's forests, the idea goes, they will lead a movement in "biocultural restoration": by becoming reenchanted with biodiversity, they will be compelled to spread their new knowledge and appreciation to their neighbors. Through renewed awareness—and renewed economic dividends—rural dwellers would be empowered and inclined to protect their biodiversity patrimony (Evans, 1999; Gámez *et al.*, 1993; Janzen, 1988).

VI. THE VALUES OF BIODIVERSITY

In Costa Rica, and in many other corners of the earth, biologists are working to raise awareness of the many values of biodiversity. Biologists hope that biodiversity's appeal to many different audiences will spark diverse efforts to preserve it. Some of the values biologists attribute to biodiversity find their loci in biodiversity itself; in most, though, humans are the value holders.

As we might expect, biologists promote the value biodiversity holds for science, as the raw material for scientists' investigations: it is "Earth's living library" (Lovejoy, 1992). Also unsurprisingly, biologists speak extensively about biodiversity's ecological value. Biodiversity provides numerous "ecosystem services" (or, depending on your definition, "ecosystem services" is part of what constitutes "biodiversity"). That is to say, biodiversity keeps global ecosystems functioning. Biodiversity purifies water, controls agricultural pests, decomposes waste, pollinates plants, stabilizes global climate, creates soil, transports nutrients, controls erosion, and maintains the ecological matrix human society requires to exist at all (Daily, 1997; Tilman, 1999). And as humans attempt to repair damaged ecosystems, we will need reservoirs of intact biodiversity to restore what we have destroyed (Jordan, 1997).

For those who measure value in terms of money, biologists assert that biodiversity has vast economic value. Biologists note that biodiversity's value to functioning ecosystems is priceless—although Costanza *et al.* (1997) have tried to calculate the incalculable. Costanza's team figured that 17 ecosystem services in 16 biomes contribute between 16 and 54 trillion (U.S) dollars per year to human economies, and they believe this is a conservative estimate. This, they note, is higher than the global gross national product of 18 trillion dollars per year. Janzen (1986) offers an exhaustive list of the treasures biodiversity already has directly provided human society. Others (e.g., Wilson, 1992) discuss the lucre that can be gained from the hidden food, medicine, chemicals, fibers, and other goods lurking in the world's natural places. As noted earlier, Costa Rica has invested heavily in its biodiversity; through ecotourism and bioprospecting, the nation is profiting from its biodiversity resources. Biologists (e.g., Janzen, 1990) also tout biodiversity's social amenity value: it can contribute to sustainable development efforts (through direct harvesting or ecotourism), foster national pride, or help those living nearby to lead more intellectually and aesthetically fulfilled lives.

E. O. Wilson (e.g., 1992, 1998) has been the leading proponent of biodiversity's value to fulfill our biophilic impulses. Wilson and others believe that love of nature has been hardwired into our genes, and we can only be truly fulfilled by satisfying that love. Those whose biophilic passions have been rekindled will also be those who work hardest for biodiversity preservation: others refer to this as biodiversity's "transformative value." Philosopher Bryan Norton (1987) explains how interactions with biodiversity can help us reconsider our consumerist impulses and make us convert to lifeways that preserve the biodiversity that provided the impetus for transformation. Biodiversity's aesthetic value—the beauty of an individual organism and its adaptations, of landscapes, of intricate ecological processes, of the sheer riot of different life forms—may effect this kind of transformation.

When biologists suggest that biodiversity may transform us, may reawaken our biophilic impulses, they still place the locus of its value in the human valuer. Some biologists take a different tack and assert that biodiversity has intrinsic value, independent of a human valuer. This concept is difficult to prove empirically and therefore difficult for a scientist to assert. Yet given the preceding definitions of "biodiversity," if biodiversity is the totality of life forms on earth, their interactions, and the processes that gave rise to them and to us, it makes it a more complicated proposition to reject the notion of biodiversity's intrinsic value out of hand. This notion takes a turn for the spiritual, and, in fact, some biologists also discuss biodiversity's spiritual value. David Ehrenfeld (1981), who would become the founding editor of *Conservation Biology*, wrote of "The Noah Principle": species "should be conserved because they exist and because this existence is itself but the present expression of a continuing historical process of immense antiquity and majesty. Long-standing existence in Nature is deemed to carry with it the unimpeachable right to continued existence." Some biologists (see quotes in Takacs, 1996) agree that the value inherent in biodiversity makes it sacred and others put the locus of value in humans, and they discuss the spiritual nourishment that contact with biodiversity brings.

The multiplicity of meanings of "biodiversity" is reflected in the multiplicity of values biologists find in it. For those biologists who would raise awareness of biodiversity's plight, it makes sense to speak for as many different values—and therefore to as many different audiences—as possible.

VII. HISTORICAL AWARENESS OF BIODIVERSITY REDUX

We can read the title of this encyclopedia entry in many ways. Smithsonian tropical biologist and biodiversity advocate Terry Erwin says (in Takacs, 1996, pp. 100–101): "I've always kind of been ahead of the game or ahead of the thinking in my own little field of entomology . . . I think there will be individuals who can influence the direction of change. And I'd like to be among them, for whatever little part I might be able to play." Ideas can act as forces of nature. They can change ecologies. They can reshape how we value and therefore how we treat nature. On behalf of biodiversity, biologists are making history: they are changing the course of events so that human history and natural history will unfurl to their liking. This encyclopedia attempts to inform, to raise awareness of both the science of biodiversity and the conscience of those who study it, revere it, and wish to see it preserved.

History suggests that two phenomena will continue to obtain: the biological complexity of the earth will continue to diminish, and biologists will continue to look for strategies that will compel us to care about what they care about and support their authority to speak for those entities. Biologists have attempted to raise our awareness of biodiversity: the complexity of real world organisms, species, and processes commingled with biologists' factual, political, emotional, ethical, aesthetic, and spiritual values of the natural world, all combined to shape public perceptions, actions, and feelings. To become aware of it is to understand not only how biologists understand the riot of life and interconnections, but to become aware of our own obligations to future evolutionary history.

See Also the Following Articles

BIODIVERSITY, DEFINITION OF • BIODIVERSITY-RICH COUNTRIES • CONSERVATION BIOLOGY, DISCIPLINE OF • CONSERVATION EFFORTS, CONTEMPORARY • DARWIN, CHARLES • ECONOMIC VALUE OF BIODIVERSITY, OVERVIEW • FOREST CANOPIES, ANIMAL DIVERSITY • STEWARDSHIP, CONCEPT OF

Bibliography

Anonymous. (1994) An interview with the father of biodiversity. *Nature Conservancy*, July–Aug: 25–29.

Bowler, P. (1993). *The Norton History of the Environmental Sciences*. Norton, New York.

Carson, R. (1962, 1987). *Silent Spring*. Houghton Mifflin, Boston.

Costanza, R. d'Arge, R., deGroot, R., Farber, S., Grasso, M., Hannon, B., Limburg, K., Naeem, S., O'Neill, R. V., Pacuelo, J., Raskin, R. G., Sutton, P., and van den Belt, M. (1997). The value of the world's ecosystem services and natural capital. *Nature* 387, 253.

Cronon, W. (1995). Introduction: In search of nature. In *Uncommon Ground: Toward Reinventing Nature* (W. Cronon, Ed.). W. W. Norton, New York.

Daily, G. C. (Ed.) (1997). *Nature's Services: Societal Dependence on Natural Ecosystems*. Island Press, Washington, DC.

Ehrenfeld, D. (1981). *The Arrogance of Humanism*. Oxford University Press, New York.

Ehrlich, P. (1985). Extinctions and ecosystem functions: Implications for humankind. In *Animal Extinctions: What Everyone Should Know* (R. J. Hoage, Ed.). Smithsonian Institution Press, Washington, DC.

Ehrlich, P. R., and Ehrlich, A. H. (1981). *Extinction: The Causes and Consequences of the Disappearance of Species*. Ballantine Books, New York.

Elton, C. S. (1958). *The Ecology of Invasions by Animals and Plants*. Methuen, London.

Erwin, T. (1991). A plan for developing consistent biotic inventories in temperate and tropical habitats. Part 1 of *Establishing a Tropical Species Co-ocurrence Database* (T. L. Erwin, Ed.). Memorias del Museo de Historia Natural, parts 1–3. Universidad Nacional Mayor de San Marcos, Lima.

Evans, S. (1999). *The Green Republic: A Conservation History of Costa Rica*. University of Texas Press, Austin.

Evernden, N. (1992) *The Social Creation of Nature*. Johns Hopkins University Press, Baltimore.

Gámez, R., Piva, A., Sittenfeld, A., Leon, E., Jimenez, J., and Mirabelli, G. (1993). Costa Rica's Conservation Program and National Biodiversity Institute. In *Biodiversity Prospecting* (W. V. Reid, S. A. Laird, C. A. Meyer, R. Gámez, A. Sittenfeld, D. H. Janzen, M. A. Gollin, and C. Juma, Eds.). World Resources Institute, Washington, DC.

Guha, R. (1989). Radical American environmentalism and wilderness preservation: A third world critique. *Environmental Ethics* 11, 71–83.

Janzen, D. H. (1986). The future of tropical ecology. *Annual Review of Ecology and Systematics* 17, 305–324.

Janzen D. H. (1988). Tropical ecological and biocultural restoration. *Science* 239, 243–244.

Janzen, D. H. (1990). "Sustainable society through applied ecology: The reinvention of the village. In *Race to Save the Tropics* (R. Goodland, Ed.). Island Press, Washington, DC.

Jordan, W. (1997) Ecological restoration and the conservation of biodiversity. In *Biodiversity II: Understanding and Protecting Our Biological Resources*. (M. L. Reaka-Kudla, D. E. Wilson, and E. O. Wilson, Eds.). Joseph Henry Press. Washington,DC.

Kellert, S. R. (1986). Social and perceptual factors in the preservation of animal species. In *The Preservation of Species: The Value of Biological Diversity* (B. Norton, Ed.). Princeton University Press, Princeton.

Kellert, S. R. (1996). *The Value of Life: Biological Diversity and Human Society*. Island Press, Washington, D.C.

Kohm, K. (Ed.) (1991). *Balancing on the Brink of Extinction: The Endangered Species Act and Lessons for the Future*. Island Press, Washington D C.

Leopold, A. (1970, 1949). *A Sand County Almanac*. Sierra Club, San Francisco.

Lovejoy, T. E. (1992). Earth's living library: Check it out. *Washington Post*, 19 March, A27.

Mann, C. C., and M. L.Plummer (1995). *Noah's choice: The future of endangered species*. Knopf, New York.

Marsh, G. P. (1864, 1965). *Man and Nature*. Charles Scribner, New York.

Myers, N. (1979) *The Sinking Ark: A New Look at the Problem of Disappearing Species*. Pergamon Press, Oxford.

Nabhan, G. P. (1997). *Cultures of Habitat: On Nature, Culture, and Story*. Counterpoint, Washington, DC.

Norton, B. (1987). *Why Preserve Natural Variety?* Princeton University Press, Princeton.

Peters, R. L., and T. E. Lovejoy (Eds.) (1992). *Global Warming and Biological Diversity*. Yale University Press, New Haven.

Reid, W., Barber, C., and Miller, K. (1992). *Global Biodiversity Strategy: Guidelines for Action to Save, Study, and Use Earth's Biotic Wealth Sustainably and Equitably*. World Resources Institute, World Conservation Union, and United Nations Development Program, Washington, DC.

Soulé, M. (1985). What is conservation biology? *BioScience* 35, 727–734.

Stork, N. E. (1997). Measuring global diversity and its decline. In *Biodiversity II: Understanding and Protecting Our Biological Resources* (M. L. Reaka-Kudla, D. E. Wilson, and E. O. Wilson, Eds.). Joseph Henry Press, Washington, DC.

Takacs, D. (1996). *The Idea of Biodiversity*. Johns Hopkins University Press, Baltimore.

Tilman, D. (1999). The ecological consequences of changes in biodiversity: A search for general principles. *Ecology* 80(5): 1455–1474.

Williams, R. (1980) Ideas of nature. In *Problems in Materialism and Culture*. Verso, London.

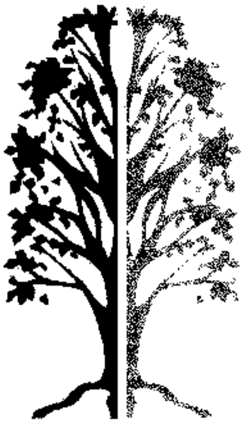
Wilson, E. O. (1987). The little things that run the world (the importance and conservation of invertebrates). *Conservation Biology* 1, 344–346.

Wilson, E. O. (Ed.) (1988). *BioDiversity*. Smithsonian, Washington, DC.

Wilson, E. O. (1992). *The Diversity of Life*. Harvard University Press, Cambridge.

Wilson, E. O. (1998). *Consilience: The Unity of Knowledge*. Alfred A. Knopf, New York.

Worster, D. (1987). *Nature's Economy: A History of Ecological Ideas*. Cambridge University Press, Cambridge.



HOTSPOTS

Norman Myers
Green College, Oxford University

GLOSSARY

biodiversity Popularly supposed to refer to the spectrum of all species on Earth, the concept should also include species' subunits (genetic diversity) and the diversity of ecosystems and ecological processes.

endemics Those species that are limited to relatively small areas, being found nowhere else on Earth.

hotspots Those areas that (a) feature exceptional concentrations of endemic species and (b) face imminent threat of habitat destruction.

megadiversity Phenomenon of at least 70% of all species being confined to 17 "megadiversity" countries.

HOTSPOTS are specific areas of the Earth's land surface that have a disproportionately large number of extant species. Identification of the world's hotspots—roughly 18–25 in number, depending on the criteria employed—provides a means of focusing on those areas where threats to biodiversity are most extreme and conservation efforts can be most effective. Underlying this approach is the thesis that present conservation resources are not sufficient to maintain all threatened species and thus global priorities need to be established.

I. INTRODUCTION

We are witnessing the opening phase of a mass extinction episode that, if allowed to persist, could well eliminate a large proportion of all species among other forms of biodiversity within the foreseeable future (Myers, 1993; Wilson, 1992). We do not have nearly enough conservation resources (funds, scientific skills, and the like) to assist all species under threat, and as the biotic crisis gathers momentum the shortfall will become ever more severe. This predicament places a premium on priority planning. Which conservation strategies offer the biggest payoff? Or, to be more precise: How can we save the most species at the least cost? This key question is likely to remain at the forefront of conservation endeavors as the Earth's biotic crisis grows worse. By concentrating on a few critical areas where needs are

Encyclopedia of Biodiversity, Volume 3

greatest and where the payoff from safeguard measures would also be greatest, conservationists can engage in a more systematized response to the challenge of large-scale extinctions that lie ahead. This represents a focused silver bullet response, in contrast to the scattergun approach that has characterized much conservation activity to date.

This is not to say—and the point is emphasized—that biodiversity outside of species-rich areas should be ignored. The biodiversity of any country is vitally important to that country's environmental well-being. A number of responses have been proposed. The "hotspots" thesis is one such mode of setting priorities at a time when we need to determine priorities with more scientific acumen than ever. This approach identifies areas that feature exceptional concentrations of endemic species and that face exceptional threat of imminent habitat destruction (Myers, 1988, 1990; Myers *et al.*, 2000). A hotspots strategy can be complemented with measures that highlight "megadiversity" countries, that is, those few countries that harbor most of the world's species, whether threatened or not (Mittermeier and Mittermeier, 1997). A further backup strategy is the protection of wilderness areas, being extensive tracts of little-disturbed wildlands with rich biodiversity stocks where conservationists can "get it right" from the start. Still a fourth response would focus on "high-value" ecosystems that, while not harboring unusual concentrations of species, encompass other remarkable manifestations of biodiversity. We shall consider each of these strategies, while giving most attention to the most promising option, the conservation of hotspots.

II. BIODIVERSITY HOTSPOTS AS ORIGINALLY IDENTIFIED

The hotspots strategy was first raised in the late 1980s, when an exploratory listing of hotspots identified 18 localities, 14 in tropical moist forests and 4 in Mediterranean-type zones (Myers, 1988, 1990). The analysis centered on higher plant species alone, with the assumption that these would serve as acceptable indicator taxa for other categories of species (mammals, birds, and other vertebrates, plus invertebrates). Analysis revealed that at least 20% of all plant species were confined to areas comprising 0.5% of Earth's land surface—areas that for the most part have already lost the bulk of their biodiversity habitats.

III. REVISED AND EXPANDED HOTSPOTS ANALYSIS

A more recent effort focusing on 25 biodiversity hotspots (Table I) has sought to refine and expand the original hotspots analysis by including other taxa and biomes (Myers *et al.*, 2000). It has centered not only on higher plants but also on birds, mammals, reptiles, and amphibians as indicators of vertebrate taxa. In addition, it has extended ecozone coverage to include tropical dry forests, woodlands, savanna and open savanna, temperate moist forests, grasslands, and arid lands.

This expanded approach still omits invertebrates from consideration, which is regrettable in that invertebrates comprise the great majority of all species, at least 95% and possibly 99%. Future analysis could be extended to include butterflies as indicators of invertebrate hotspots. Butterflies are more closely tied to plant communities than vertebrates, and yet are popular enough with amateur naturalists that for some areas we have quite accurate records of their populations through time. (As a bonus factor, butterflies are sometimes the best single group of animals as indicators of ecosystem health.) Additional support for invertebrate hotspots could be gained from focusing on dragonflies, damselflies, and tiger beetles, all of which are widespread and fairly well known.

This latest hotspots analysis, like the earlier one, is limited to terrestrial biotas. It is also based on vascular or higher plants (which comprise around 90% of all plants, and are hereafter referred to as "plants"), since they are essential to virtually all forms of animal life. These plants are well known, with their conservation status adequately documented for the most part.

The endemism data tend to be minimal for two reasons. One is the sheer lack of recent documentation in the form of, for example, modern floras. For instance, there is no up-to-date account of Brazil's plant species even though the country is believed to harbor 50,000 species or one-sixth of the world's total. More importantly, endemism statistics almost always relate only to individual countries or parts of countries, whereas 12 of the hotspots extend across two or more countries and 6 hotspots across four or more. In these cases, it has been singularly difficult to compute totals for hotspot-wide endemics with cross-boundary complications, and the analysis (Myers *et al.*, 2000) has usually had to depend on best-judgment estimates by scientists with extensive on-ground experience. These estimates tend to be cautious and conservative, meaning that the

TABLE I

The 25 Biodiversity Hotspots[a]

Hotspot	Original extent (km²)	Remaining primary vegetation (km²) (% of original extent)	Area protected (km²) (% of hotspot)	Plant species	Endemic plants (% of global plants total, 300,000)	Vertebrate species	Endemic vertebrates (% of global vertebrates total, 27,298)
Tropical Andes	1,258,000	314,500 (25.0)	79,687 (25.3)	45,000	20,000 (6.7%)	3389	1567 (5.7%)
Mesoamerica	1,155,000	231,000 (20.0)	138,437 (59.9)	24,000	5000 (1.7%)	2859	1159 (4.2%)
Caribbean	263,500	29,840 (11.3)	29,840 (100.0)	12,000	7000 (2.3%)	1518	779 (2.9%)
Brazil's Atlantic Forest	1,227,600	91,930 (7.5)	33,084 (35.9)	20,000	8000 (2.7%)	1361	567 (2.1%)
Choco/Darien/Western Ecuador	260,600	63,000 (24.2)	16,471 (26.1)	9000	2250 (0.8%)	1625	418 (1.5%)
Brazil's Cerrado	1,783,200	356,630 (20.0)	22,000 (6.2)	10,000	4400 (1.5%)	1268	117 (0.4%)
Central Chile	300,000	90,000 (30.0)	9167 (10.2)	3429	1605 (0.5%)	335	61 (0.2%)
California Floristic Province	324,000	80,000 (24.7)	31,443 (39.3)	4426	2125 (0.7%)	584	71 (0.3%)
Madagascar[b]	594,150	59,038 (9.9)	11,548 (19.6)	12,000	9704 (3.2%)	987	771 (2.8%)
Eastern Arc and Coastal Forests of Kenya/Tanzania	30,000	2000 (6.7)	2000 (100.0)	4000	1500 (0.5%)	1019	121 (0.4%)
Guinean Forests of West Africa	1,265,000	126,500 (10.0)	20,324 (16.1)	9000	2250 (0.8%)	1320	270 (1.0%)
Cape Floristic Province	74,000	18,000 (24.3)	14,060 (78.1)	8200	5682 (1.9%)	562	53 (0.2%)
Succulent Karoo	112,000	30,000 (26.8)	2352 (7.8)	4849	1940 (0.6%)	472	45 (0.2%)
Mediterranean Basin	2,362,000	110,000 (4.7)	42,123 (38.3)	25,000	13,000 (4.3%)	770	235 (0.9%)
Caucasus	500,000	50,000 (10.0)	14,050 (28.1)	6300	1600 (0.5%)	632	59 (0.2%)
Sundaland	1,600,000	125,000 (7.8)	90,000 (72.0)	25,000	15,000 (5.0%)	1800	701 (2.6%)
Wallacea	347,000	52,020 (15.0)	20,415 (39.2)	10,000	1500 (0.5%)	1142	529 (1.9%)
Philippines	300,800	9023 (3.0)	3910 (43.3)	7620	5832 (1.9%)	1093	518 (1.9%)
Indo-Burma	2,060,000	100,000 (4.9)	100,000 (100.0)	13,500	7000 (2.3%)	2185	528 (1.9%)
South-Central China	800,000	64,000 (8.0)	16,562 (25.9)	12,000	3500 (1.2%)	1141	178 (0.7%)
Western Ghats/Sri Lanka	182,500	12,450 (6.8)	12,450 (100.0)	4780	2180 (0.7%)	1073	355 (1.3%)
SW Australia	309,850	33,336 (10.8)	33,336 (100.0)	5469	4331 (1.4%)	456	100 (0.4%)
New Caledonia	18,600	5200 (28.0)	526.7 (10.1)	3332	2551 (0.9%)	190	84 (0.3%)
New Zealand	270,500	59,400 (22.0)	52,068 (87.7)	2300	1865 (0.6%)	217	136 (0.5%)
Polynesia/Micronesia	46,000	10,024 (21.8)	4913 (49.0)	6557	3334 (1.1%)	342	223 (0.8%)
Total	17,444,300	2,122,891 (12.2)	800,767 (37.7)	**[c]	133,149 (44%)	**[c]	9645 (35%)

[a] Choco/Darien/Western Ecuador stretches from the Darien of Panama to Western Ecuador; Cape Floristic Province lies in the southern sector of South Africa; the Succulent Karoo lies in western South Africa; Sundaland encompasses the islands of western Indonesia together with the Malay Peninsula; and Wallacea includes islands of eastern Indonesia.

[b] Madagascar includes the nearby islands of Mauritius, Reunion, Seychelles, and Comores.

[c] These totals cannot be summed due to overlapping between hotspots.

Source: Myers, *et al.*, 2000.

true totals will likely be higher than those presented here.

To qualify as a hotspot, the main determining criterion is species endemism. A second criterion is degree of threat; to qualify, an area must retain only 30% or less of its original primary vegetation. The cutoff adopted is 1500 endemic plant species, or 0.5% of the 300,000 plant species on land (Prance *et al.*, 2000).

Unlike the earlier hotspots analysis, the expanded version includes birds, mammals, reptiles, and amphibians. It excludes the only other vertebrate group, fishes (totaling roughly half of all vertebrates), because data about their numbers, habitats, and conservation status are generally poor. Hereafter, the term "vertebrates" refers to all vertebrates except fishes. Not that vertebrates serve as a second determinant of hotspot status; if an area qualifies by the 0.5% plants criterion, it makes the list. Vertebrates are strictly used for backup support and to determine measures of congruence. Sixteen hotspots contain at least 130 vertebrate endemics, or 0.5% of the 27,298 species of the four vertebrate groups worldwide, while 12 of these contain at least twice as many.

As noted, the analysis omits insects and other invertebrates. While scientists have documented the great majority of plants and vertebrates, they have documented only a tiny proportion of invertebrates. For present purposes and purely as a working proposition, it is reasonable to assume that the five categories of endemic species documented here are roughly matched by similar concentrations of endemic invertebrate species. Although this assumption is preliminary and approximate, it is supported by more evidence in its favor than against.

Many of the hotspots have already lost 90% of their original primary vegetation, and a few of them, for example, the Mediterranean Basin, Indo-Burma, and the Philippines, have lost at least 95%. It is true that disrupted and secondary vegetation can sometimes support moderate numbers of original species, but in the main this is not significant for conservation purposes.

Constraints of socioeconomic status, political commitment, or conservation feasibility in the countries concerned have not been considered. This is partly because these factors are difficult to quantify, and partly because they can be better incorporated when designing conservation projects. All the same, it is worth noting that Nepal, for instance, hardly possesses the administrative structures, the managerial know-how, or the planning capacities that can usefully absorb additional external assistance. Related questions formerly arose with respect to Zaire (now the Democratic Republic of Congo): Should that country have been granted conservation support when its military spending was twice as large a proportion of its gross domestic product as the average for sub-Saharan Africa, and when the personal wealth of President Mobutu was greater than the entire country's economy?

IV. MAIN FINDINGS OF REVISED HOTSPOTS ANALYSIS

A total of 25 hotspots contain the remaining habitats of 133,149 plant species (44% of all plant species) and of 9645 vertebrate species (35% of all such species). These endemics are confined to an aggregate area equivalent to 1.4% of Earth's land surface (Table I). The hotspots are so threatened that, having already lost at least 70% of their original primary vegetation, they all seem likely, in the absence of greatly increased conservation efforts, to lose much if not most of their remaining vegetation within the foreseeable future.

The 25 hotspots feature a broad range of ecosystem types. Predominant are tropical rain forests (in 15) and Mediterranean-type zones (in 5). Nine are mainly or completely made up of islands; almost all tropical islands fall into one or another hotspot. Sixteen hotspots are in the tropics, which largely places them in developing countries where threats are greatest and conservation resources are in shortest supply.

Now consider the relationship of endemic species to total species. In the 17 "megadiversity" countries with some 70% of Earth's species, the ratio of endemic nonfish vertebrates to all vertebrates ranges from a high of 1 : 1.3 for Madagascar to a low of 1 : 14 for the Democratic Republic of Congo (formerly Zaire). When all 25 hotspots are considered, the average ratio is 1 : 2.8. This high ratio of endemism demonstrates the significance and rarity of the biodiversity found in these hotspots.

Still more significant, the extent of habitat loss in the hotspots means that we can reasonably assume they harbor an even greater share of threatened species, defined here as Red Data Book species (these species are assessed by the World Conservation Union and include only species known to science and known to be threatened; the true total is far higher). So far as we can calculate, albeit in a preliminary and exploratory manner, the number of threatened species occurring in hotspots probably amounts to roughly two-thirds of all threatened species.

A. The "Hotter" Hotspots

Some hotspots are "hotter" than others. In nine hotspots, 30% of all plants are endemics (in the Tropical Andes, an exceptional 6.7% are endemic) and 25% of vertebrate species are endemics; these hotspots account for 0.7% of Earth's land surface (Table II). At the same time, they feature some of the most depleted habitats anywhere: Madagascar retains less than 10% of its original primary vegetation; Sundaland and Brazil's Atlantic Forest less than 8%; the Mediterranean Basin, Indo-Burma, and the Philippines less than 5%. Five hotspots hold more than 2% of the world's biodiversity in both plants and vertebrates, hence they are super hotspots: Tropical Andes, 6.7% and 5.7%, respectively; Sunda-

TABLE II

Leading Hotspots in Terms of Endemic Species

Hotspot	Endemic plants (% of global plants total, 300,000)	Endemic vertebrates (% of global vertebrates total, 27.298)
Tropical Andes[a]	20,000 (6.7)	1567 (5.7)
Sundaland[a]	15,000 (5.0)	701 (2.6)
Madagascar[a]	9704 (3.2)	771 (2.8)
Brazil's Atlantic Forest[a]	8000 (2.7)	567 (2.1)
Caribbean[a]	7000 (2.3)	779 (2.9)
Subtotal	59,704 (19.9)	4385 (16.1)
Mesoamerica	5000 (1.7)	1159 (4.2)
Mediterranean Basin	13,000 (4.3)	235 (0.9)
Indo-Burma	7000 (2.3)	528 (1.9)
Philippines	5832 (1.9)	519 (1.9)
Total	90,536 (30.1)[b]	6826 (25.0)

[a] Hotspots with at least 2% of both endemic plants and vertebrates. and together comprising only 0.4% of Earth's land surface (all nine hotspots amount to 0.7% of Earth's land surface).

[b] This would total 30.2% but for rounding of numbers in the individual hotspots.

land, 5.0% and 2.6%; Madagascar, 3.2% and 2.8%; Brazil's Atlantic Forest, 2.7% and 2.1%; and the Caribbean, 2.3% and 2.9%. Collectively these five areas account for 20% of all endemic plant species and 16% of all endemic vertebrate species in just 0.4% of Earth's land surface. They also harbor 45% of the plant and vertebrate endemics in the 25 identified hotspots.

B. Species/Area Relationships

Some hotspots are also significant because their endemic species are concentrated in exceptionally small areas (Table III). The Eastern Arc and Coastal Forests of Tanzania/Kenya (hereafter referred to as "Eastern Arc") contain 1500 endemic plants in 2000 km^2 for a ratio of 75 species to 100 km^2, or 75 : 1, and 121 endemic vertebrates for a ratio of 6 : 1—both ratios top the lists for all hotspots. New Caledonia (5200 km^2) has ratios of 49 : 1 and 1.6 : 1 for endemic plants and vertebrates, the Philippines (9023 km^2) has 64.7 : 1 and 5.7 : 1, Polynesia/Micronesia (10,024 km^2) has 33 : 1 and 2.2 : 1, and the Western Ghats in India (12,450 km^2) has 17.5 : 1 and 2.9 : 1. Ratios for the other areas range from 18 : 1 to 1.2 : 1 for plants and from 2.9 : 1 to 0.03 : 1 for vertebrates.

TABLE III

Species/Area Ratios per 100 km^2 of Hotspots

Hotspot	Endemic plants	Endemic vertebrates
Tropical Andes	6.4	0.5
Mesoamerica	2.2	0.5
Caribbean	23.5	2.6
Brazil's Atlantic Forest	8.7	0.6
Choco/Darien/Western Ecuador	3.6	0.7
Brazil's Cerrado	1.2	0.03
Central Chile	1.8	0.06
California Floristic Province	2.7	0.09
Madagascar	16.4	1.3
Eastern Arc and Coastal Forests of Kenya/Tanzania	75.0	6.1
Guinean Forests of West Africa	1.8	0.2
Cape Floristic Province	31.6	0.3
Succulent Karoo	6.5	0.15
Mediterranean Basin	11.8	0.2
Caucasus	3.2	0.1
Sundaland	12.0	0.6
Wallacea	2.9	1.0
Philippines	64.7	5.7
Indo-Burma	7.0	0.5
South-Central China	5.5	0.3
Western Ghats/Sri Lanka	17.5	2.9
SW Australia	13.0	0.3
New Caledonia	49.1	1.6
New Zealand	3.1	0.2
Polynesia/Micronesia	33.3	2.2

C. Congruence among Species Categories

In several hotspots, there is a measure of congruence between plants and vertebrates insofar as high counts for endemic plants are matched by moderately high counts for endemic vertebrates (Table IV). This factor can reinforce the conservation priority thesis, especially in those hotspots with the most endemic species. There can also be high congruence in areas with lower species counts, for example, there is 100% congruence in the Philippines with 1.9% of both endemic plants and vertebrate species worldwide, and 80% congruence in the Eastern Arc with 0.5% of plant species and 0.4% of vertebrate species. But the species percentages of these areas are low relative to those of several other hotspots.

The Tropical Andes holds 6.7% of all endemic plant species worldwide and 5.7% of endemic vertebrates, for 85% congruence. Madagascar's endemic species repre-

TABLE IV

Congruence between Endemic Plants and Vertebrates

Hotspot	Endemic plants as % of global plants total, 300,000	Endemic vertebrates as % of global vertebrates total, 27,298	% Congruence (rounded)
Tropical Andes	6.7%	5.7%	85
Mesoamerica	1.7%	4.2%	41
Caribbean	2.3%	2.9%	79
Brazil's Atlantic Forest	2.7%	2.1%	78
Choco/Darien/Western Ecuador	0.8%	1.5%	53
Brazil's Cerrado	1.5%	0.4%	27
Central Chile	0.5%	0.2%	40
California Floristic Province	0.7%	0.3%	43
Madagascar	3.2%	2.8%	88
Eastern Arc and Coastal Forests of Kenya/Tanzania	0.5%	0.4%	80
Guinean Forests of West Africa	0.8%	1.0%	80
Cape Floristic Province	1.9%	0.2%	11
Succulent Karoo	0.6%	0.2%	33
Mediterranean Basin	4.3%	0.9%	21
Caucasus	0.5%	0.2%	40
Sundaland	5.0%	2.6%	52
Wallacea	0.5%	1.9%	26
Philippines	1.9%	1.9%	100
Indo-Burma	2.3%	1.9%	83
South-Central China	1.2%	0.7%	58
Western Ghats/Sri Lanka	0.7%	1.3%	54
SW Australia	1.4%	0.4%	29
New Caledonia	0.9%	0.3%	33
New Zealand	0.6%	0.5%	83
Polynesia/Micronesia	1.1%	0.8%	73

sent 3.2% and 2.8%, for 88% congruence; the Caribbean has 2.3% and 2.9%, to give 79% congruence. (The Tropical Andes is a large area where one could expect high congruence; the other two are only one-fifth and one-tenth as big respectively.) By contrast, the Cape Floristic Province possesses 1.9% of all endemic plants but only 0.2% of all endemic vertebrates. Congruence tends to be high in tropical forest hotspots, and generally low in Mediterranean-type hotspots (the congruence for Cape Floristic Province is 11%) and other drier areas with their meager counts for endemic vertebrates.

The four vertebrate groups reveal varying degrees of congruence among themselves. Birds and amphibians (like plants) generally show an increase in species numbers in the tropics and still more nearer the equator, with particularly high totals in tropical forests. Their numbers also increase with altitude up to 2500 m in localities with good rainfall. Similarly, reptile abundance increases nearer the equator, though in drier zones it is comparable to that in tropical forests, as witness the situation in the two countries with the most reptiles, Australia and Mexico. Mammal numbers also increase closer to the equator, with drier areas again having a species richness comparable to that of tropical forests. Certain groups within the mammals, notably primates and bats, show similar trends to the birds and amphibians, though primates reveal their greatest species numbers in lowland rain forests, declining rapidly with altitude.

D. The "Hottest" Hotspots

It is not practicable to devise a hotspots ranking index that combines five criteria, namely, numbers of endemies and endemic species/area ratios for both plants and vertebrates, and habitat loss. These criteria cannot carry equal weight (a case of comparing apples and oranges), so one cannot simply sum the rankings for each case. For comparative purposes, Table V lists the eight "hottest hotspots" that appear at least three times in the top ten rankings for each criterion. The leaders are Madagascar, the Philippines, and Sundaland, which appear in all five criteria, followed by the Caribbean and Brazil's Atlantic Forest, which appear in four. Three of these hotspots (Madagascar, the Philippines, and the Caribbean) have small land areas, which further highlights their importance.

Two additional hotspots, the Tropical Andes and Mediterranean Basin, should be considered as candidates for conservation support in light of their exceptional totals for endemic plants (Tropical Andes ranks highest for both endemic plants and vertebrates, and the Mediterranean ranks third for endemic plants). Yet they do not appear in Table V because they ranked in the top ten in only two criteria listings. Similarly, Mesoamerica ranks second for endemic vertebrates (49% higher than the third-ranking Caribbean), but scores only tenth for endemic plants.

E. Higher Taxa Assessment

Quantitative analysis can be complemented by a qualitative evaluation of endemism among higher taxa such as families and genera. Yet sufficient sampling and exhaustive surveys are not available to measure the distribution of biodiversity overall. A top-down taxonomic approach, however, could compare the biodiversity of different areas using measures based on the number of higher taxa in each. For instance, family richness can often be a good predictor of species richness for certain groups and regions, including British ferns and butterflies, Australian passerine birds, and North and Central American bats, indicating that higher taxa indicators may sometimes offer a valid shortcut assessment.

Madagascar possesses 11 endemic families and 310 endemic genera of plants, 5 endemic families and 14 endemic genera of primates, and 5 endemic families and 35 endemic genera of birds. The Cape Floristic Province has 6 endemic families and 198 endemic genera of plants; and New Caledonia has 5 endemic families and 112 endemic genera of plants, plus 1 endemic family and 3 endemic genera of birds. By contrast, the United States and Canada, with an area 8.8 times larger than that of the 25 hotspots combined, have only 2 endemic families of plants. Plant family richness can often serve as a predictor of species richness for certain animal groups, such as mammals, amphibians, and reptiles.

F. Action Responses

To review, in just 1.4% of Earth's land surface, there are 25 hotspots containing 44% of plant species and 35% of vertebrate species facing high risk of extinction. It is often estimated that, were the present mass extinction of species to proceed virtually unchecked, some-

TABLE V

The Eight "Hottest" Hotspots in Terms of Five Factors (numbers in parentheses indicate the ranking in the top 10 hotspots for each factor)

Hotspot	Endemic plants	Endemic vertebrates	Endemic plants/ area ratio (species per 100 km^2)	Endemic vertebrates/ area ratio (species per 100 km^2)	Remaining primary vegetation as % of original extent	Times appearing in top 10 for each of five factors
Madagascar	9704 (4)	771 (4)	16.4 (8)	1.3 (7)	9.9 (9)	5
Philippines	5832 (8)	518 (9)	64.7 (2)	5.7 (2)	3.0 (1)	5
Sundaland	15,000 (2)	701 (5)	12.0 (10)	0.6 (10)	7.8 (7)	5
Caribbean	7000 (6)	779 (3)	23.5 (6)	2.6 (4)	11.3	4
Brazil's Atlantic Forest	8000 (5)	567 (6)	8.7	0.6 (10)	7.5 (6)	4
Eastern Arc and Coastal Forests of Kenya/ Tanzania	1500	121	75.0 (1)	6.1 (1)	6.7 (4)	3
Indo-Burma	7000 (6)	528 (8)	7.0	0.5	4.9 (3)	3
Western Ghats/Sri Lanka	2180	355	17.5 (7)	2.9 (3)	6.8 (5)	3

where between one-third and two-thirds of all species could be eliminated within the foreseeable future. The hotspots analysis indicates that much of this threat could be countered through protection of the 25 identified hotspots.

An aggregate area of 800,767 km^2, 38% of the hotspots total, is now protected in parks and reserves. Some of these are little better than "paper parks" in offering a modicum of legal status. All are in urgent need of stronger safeguards, including those five hotspots that fall entirely within protected areas. The areas without any protection at all amount to 1.3 million km^2 or 62% of the hotspots total. In a few areas, new safeguards will not provide outright protection of a traditional sort because human settlements and other activities are well established. These areas could receive a measure of protection as "conservation units" that allow some degree of multiple use provided that species safeguards are always paramount. In short, the prospect of a mass extinction can be diminished and conservation efforts can be more effective by applying a hotspots strategy.

The hotspots findings reported here complement several other priority-setting analyses. There is a 68% overlap with Birdlife International's Endemic Bird Areas (Stattersfield *et al.*, 1998), 82% with the World Conservation Union (IUCN)/World Wide Fund for Nature International Centres of Plant Diversity and Endemism (Davis *et al.*, 1994–1997), and 92% with the most critical and endangered ecoregions of the World Wildlife Fund–US Global 200 List (Dinerstein *et al.*, 1996). The hotspots approach is more comprehensive than the first two by combining five categories of species, and is more tightly focused than the third.

There are surely other hotspots that feature exceptional plant endemism and face serious threat but that are not sufficiently documented to meet the hotspots criteria. They include the Ethiopian Highlands, the Angola Escarpment, southeastern China, Taiwan, and the forests of the Albertine Rift in eastern Democratic Republic of Congo (formerly Zaire), southwestern Uganda, and northern Rwanda. Were these five areas to be added to the hotspots list, they would increase the plants endemics total by only a few percent.

In addition, there are a good many mini-hotspots. One such is Queensland's Wet Tropics and adjacent tropical forest tracts along the Queensland coast, which contains a host of endemic species, with an exceptionally high species/area ratio (around 1200 endemic plants in less than 11,000 km^2).

To reiterate a key point: the biodiversity hotspots are not the only areas that deserve priority treatment from conservation planners. Indeed, every country has its own biodiversity stocks, even if they are not as diverse or as concentrated as those of the major hotspots. We shall now look at three other criteria that warrant attention for conservation priority rankings.

V. WILDERNESS AREAS

There are a few tropical forest expanses known as "major tropical wilderness areas" (Mittermeier *et al.*, 1998) or "good news" areas (Myers, 1988, 1990) that total some 6–7 million km^2 and feature concentrations of endemic species while retaining at least 75% of their primary vegetation. These areas also have fewer than five persons per square kilometer.

One is the island of New Guinea, which has around 15,000 endemic plants. Others include the Guayana Shield of northeastern Amazonia, the lowlands of western Amazonia, and the Congolian Forest, with a total of perhaps another 30,000 endemic plants. Were these regions part of a supplementary conservation strategy, they could increase the endemic plants total to almost 60% of all plant species in roughly 5% of Earth's land surface.

Wilderness areas of all kinds, that is, not just biodiversity-rich areas, comprise nearly 90 million km^2 of little-disturbed land, or well over half of Earth's land expanse. But when areas of rock, ice, and otherwise barren land are excluded, nearly three-quarters of all habitable land has been disturbed to a significant extent. It is unlikely that most wilderness areas in question will be settled by large human communities within the foreseeable future because of unfavorable climate, difficult terrain, remoteness from markets, and other factors that mean they can be readily conserved in a wilderness state. Thus these areas merit priority treatment from conservation planners. There is still opportunity in these wilderness areas for conservationists to get things right—notably in terms of land use planning—from the very start.

VI. HIGH-VALUE ECOSYSTEMS

Still another conservation strategy to be pursued in association with hotspots is to protect ecosystems of high value by reason of their exceptional abundance and concentrations of wildlife. These ecosystems contain large numbers of individual animals and large stocks of remarkable plants striking in appearance even though they comprise few species and few if any endemic species. A notable example is the Serengeti/Mara ecosystem in northern Tanzania and southern Kenya, which has 4 million wildebeest, zebras, gazelles, and other large herbivores in an area of only 25,000 km^2 (all of these species are widespread elsewhere, though not in such extraordinary numbers). For comparison, the United States has roughly 20 million deer, moose, elk, pronghorn, caribou, and other large herbivores on 9 million km^2.

Other high-value ecosystems contain exceptional concentrations of endemic species, but only in a few categories. For instance, Lake Baikal has numerous endemic fish species but few species of other sorts, whether endemic or not, and Lakes Nakuru and Naivasha in Kenya possess exceptional numbers of bird species but few endemics, and has few other species of other types. Leading candidates among these high-value ecosystems include:

- the Serengeti/Mara ecosystem;
- the monarch butterfly overwinterning sites in Mexico;
- the coastal Sundarbans area of India and Bangladesh, which has the largest tiger population left in the wild;
- Lake Baikal, with 2000-plus fish species (15% of all freshwater fishes), 1500 of them endemic;
- the East African Rift Valley lakes with 1200 fish species, 930 of which are endemic;
- Lakes Nakuru and Naivasha, which have 400-plus bird species each, and a joint total of 600 (compare to the United States total of 770); Lake Nakuru usually has 250,000 flamingoes and occasionally 2 million;
- the green turtle nesting grounds on Ascension Island in the Atlantic;
- the caribou migration lands in Alaska;
- the Galapagos Islands with their "museum of evolution";
- the Great Barrier Reef with its outstanding coral reefs; and
- the California, ecosystems that harbor giant redwoods and sequoias.

VII. MEGADIVERSITY COUNTRIES

To effectively conserve a hotspot area, it is usually critical that the country's government be committed to the conservation effort. As has been well said, hotspots do not have governments, only countries do. So a complementary approach to hotspot protection should focus on "megadiversity" countries. Such a country is defined as one that either (a) contains 20,000 higher plant species or, in the case of a country with fewer than 20,000 but more than 10,000 such species, at least 5000 endemics; or (b) contains at least 2000 species of higher vertebrates (mammals and birds), or 200 such species as endemics.

These 17 megadiversity countries encompass 60–70% of all global biodiversity (Mittermeier and Mittermeier, 1997) (Table VI). When these countries are assessed for their rankings in terms of plants, vertebrates (including freshwater fish), butterflies, and tiger beetles, the top three countries are in a class of their own, namely, Brazil, Indonesia, and Colombia. They are followed by a second group that includes Mexico, Australia, Madagascar, and Peru, and then a third group of China, the Philippines, India, Ecuador, and Venezuela. Clearly there is need to give priority attention to these 17 megadiversity countries as well as to the hotspots—though in many instances, the two lists overlap.

VIII. THE CONSERVATION IMPACT OF HOTSPOTS

The original hotspots strategy was first implemented in 1989 by the MacArthur Foundation with substantial funding for hotspots. Since then, over $400 million has been invested by international agencies and conservation groups. Yet this is only 0.8% of the total amount spent by governments during the same period on biodiversity conservation, (roughly $40 billion) and by international groups ($10 billion). These monies have been spent on across-the-board activities rather than the tightly targeted efforts advocated here. This $400 million is almost twice as much as the cost of the *Pathfinder* mission to Mars, which along with many other space probes has been justified largely on biodiversity grounds, namely, the search for extraterrestrial life.

The hotspots could be adequately protected, and thus a large proportion of all species at risk, for just $20 million per hotspot per year (12.5 times the annual average over the past ten years). The traditional scattergun approach of much conservation activity, that is, seeking to be many things to many threatened species, needs to be complemented by a well-directed hotspots strategy that emphasizes the most cost-effective measures.

Such a tightly targeted strategy could generate a handsome payoff in stemming the biotic holocaust that is now under way. In the 25 identified hotspots, 35% of Earth's nonfish vertebrate species and 44%

TABLE VI

Megadiversity Countries: Plant Diversity and Endemism

Country	Area (km^2)	Total species	Endemics
Brazil	8,511,965	~50,000–56,000	16,500–18,500
Indonesia	1,916,600	~37,000	14,800–18,500
Colombia	1,141,748	45,000–51,000	15,000–17,000
Mexico	1,972,544	18,000–30,000	10,000–15,000
Australia	7,686,810	15,638	14,458
Madagascar	587,045	11,000–12,000	8,800–9,600
China	9,561,000	27,100–30,000	~10,000
Philippines	300,780	8,000–12,000	3,800–6,000
India	3,287,782	>17,000	7,025–7,875
Peru	1,285,210	18,000–20,000	5,356
Papua New Guinea	475,369	15,000–21,000	10,500–16,000
Ecuador	283,561	17,600–21,100	4,000–5,000
USA	9,372,143	18,956	4,036
Venezuela	912,050	15,000–21,070	5,000–8,000
Malaysia	329,749	15,000	6,500–8,000
South Africa	1,221,037	23,420	16,500
Dem. Rep. Congo/Zaire	2,344,000	11,000	3,200
Total	51,189,393		155,475–183,025

of plant species currently face unusually high risk of extinction. The hotspots analysis indicates that nearly half of the overall problem could be countered through protection of the 25 hotspots, covering an aggregate area the size of three Texases or one Mexico. In short, the likelihood of a mass extinction could be greatly reduced and made much more manageable. Leading players among the "global community" are international funding organizations (e.g., the World Bank and other multilateral banks, United Nations agencies, and bilateral aid agencies), international nongovernmental organizations (e.g., conservation bodies and private foundations), and business enterprises interested in biodiversity protection (e.g., pharmaceutical corporations). All of humanity has a stake in Earth's biodiversity, and through a coordinated effort and directed actions we could make all contributions have maximum conservation impact.

Recall that the mass extinction of species, if allowed to persist, would constitute a problem far graver than all other environmental problems. We could clean up acid rain, turn back deserts, and repair the ozone layer within a matter of decades, regrow forests and restore topsoil within a century or so, and even stabilize the global climate within a few centuries. But according to evidence from mass extinctions in the prehistoric past, evolutionary processes are not likely to generate a replacement stock of species in less than 5 million years and possibly several times longer. Within the next few decades, we shall determine the future of a key feature of our biosphere, its abundance and diversity of species. The hotspots strategy offers a way to largely avoid an impoverishment of the Earth that could last at least 20 times longer than *Homo sapiens* has been a species.

Obviously, a mass extinction will also affect a large number of people. Suppose the average global population during the 5-million-year recovery period is 2.5 billion rather than the 6 billion we have today, on the grounds that the world's future ecosystems will be unable to support the current global population. This means that the total number of people who will be affected will be on the order of 500 trillion individuals. This figure dwarfs the 50 billion people who have existed so far. Even one trillion is a large number. To put it in perspective, consider the length of time made up of 1 trillion seconds

This, then, is the ultimate significance of the biotic holocaust that is overtaking the planet. Fortunately, we still have time to stem and slow the biodiversity crisis—and there are few ways to do it so successfully as by safeguarding the 25 hotspots that now harbor what nature has produced in its most exuberant expressions of life's abundance and variety.

See Also the Following Articles

BIOGEOGRAPHY, OVERVIEW • DIVERSITY, COMMUNITY/GLOBAL LEVEL • ENDEMISM • INVERTEBRATES, TERRESTRIAL, OVERVIEW • SPECIES AREA RELATIONSHIPS • HUMAN EFFECTS ON ECOSYSTEMS, OVERVIEW • VERTEBRATES, OVERVIEW

Bibliography

Davis, S. D., *et al.* (1994–1997). *Centres of Plant Diversity: A Guide and Strategy for Their Conservation* (3 vols.). World Wide Fund for Nature and International Union for Conservation of Nature and Natural Resources, Gland, Switzerland.

Dinerstein, E., *et al.* (1996). *The Global 200: Key Ecoregions for Saving Life on Earth.* World Wildlife Fund–US, Washington D. C.

Groombridge, B. (1992). *Global Biodiversity.* Chapman and Hall, London.

McNeely, J. A. (1996). *Assessing Methods for Setting Conservation Priorities.* International Union for Conservation of Nature and Natural Resources. Gland, Switzerland.

Mittermeier, R., and C. Mittermeier. (1997). *Megadiversity: Earth's Biologically Wealthiest Nations.* CEMEX, Mexico City.

Mittermeier, R. A., Myers. N., Gil, P. R., and Mittermeier, C. G. (1999). Hot spots: Earth's biologically richest and most endangered terrestrial ecosystems. CEMEX, Monterrey, Mexico, and Conservation International, Washington D.C.

Mittermeier, R. A., Myers, N., Thomsen, J. B., da Fonseca, YAB, and Oliveri, S. (1998). Biodiversity hotspots and major tropical wilderness areas: Approaches to setting conservation priorities. *Conservation Biol.* 12, 516–520.

Myers, N. (1988). Threatened biotas: "Hot spots" in tropical forests. *The Environmentalist* 8(3), 187–208.

Myers, N. (1990). The biodiversity challenge: Expanded hot-spots analysis. *The Environmentalist* 10(4), 243–256.

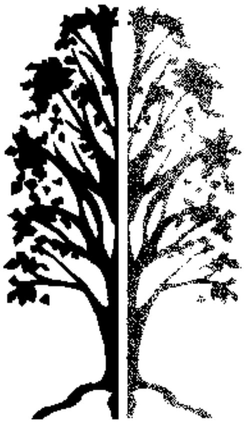
Myers, N. (1993). Questions of mass extinction. *Biodiversity and Conservation* 2, 2–17.

Myers, N., Mittermeier, R. A., Mittermeier, C. G., da Fonseca, Y. AB., and Kent, J. (2000). Biodiversity hotspots for conservation frontier. *Nature* 403, 853–858.

Prance, G. T., Beent, J. H., Dransfield, J., and Johns, R. (2000). The tropical flora remains undercollected. *Ann. Missouri Botanical Garden,* in press.

Stattersfield, A. J., Crosby, M. J., Long, A. J., and Wege, D. C. (1998). *Endemic Bird Areas of the World.* Birdlife International, Cambridge, United Kingdom.

Wilson, E. O. (1992). *The Diversity of Life.* Belknap Press, Cambridge, Massachusetts.



HUMAN EFFECTS ON ECOSYSTEMS, OVERVIEW

Paul R. Ehrlich and Claire Kremen
Stanford University

GLOSSARY

ecosystem Community of organisms in a defined area combined with the physical factors of the environment with which the community interacts.

ecosystem goods Commodities, such as timber and seafood, supplied free to humanity by natural ecosystems.

ecosystem services Functions, such as flood control and pollination, supplied free to humanity by natural ecosystems.

TWO CENTRAL ISSUES CONCERNING BIODIVERSITY TODAY are the roles that plants, animals, fungi, and microorganisms play in the functioning of ecosystems, and the primarily negative impacts of human beings on those ecosystems. Here we focus on the latter, starting with the rise of humanity as a geophysical force and then examining in some detail current anthropogenic alteration of ecosystems and the prospects of further damage by human beings to the delivery of ecosystem goods and services. Finally, we look at ways in which human damage to ecosystems could be limited in the future.

I. INTRODUCTION

Ecosystems are the functional units composed of the elements of biodiversity (plants, animals, fungi, and microorganisms) in an area interacting with each other and their nonliving environments. Ecosystems are the "natural capital" that generates a flow of ecosystem goods and services that are essential to civilization. Natural ecosystems have been the source of all foods consumed by human beings, a large proportion of the medicines, and many important industrial products ranging from natural dyes to cotton and timber. In addition to these ecosystem goods, ecosystems also supply a wide array of ecosystem services. Among other services, ecosystems tend to buffer against sudden changes in the climate, control the hydrological cycle that brings us fresh water, control floods, generate and maintain fertile soils, dispose of wastes and recycle nutrients, pollinate crops, and control the vast majority of pests that would attack crops or carry disease to human beings.

Human impacts on ecosystems precede the dawn of history. Discernible impacts probably began when people (quite possibly *Homo erectus*) first began to use fire as a tool in hunting, but there are almost insurmountable difficulties in determining when that was—

Encyclopedia of Biodiversity, Volume 3

perhaps more than a million years ago, perhaps as recently as 50,000 years ago. The skill of modern *Homo sapiens* as hunters is attested to by their probable role in the extermination of the so-called megafauna of the Pleistocene epoch (ice ages starting about a million years ago)—mammoths, cave bears, giant sloths, and the like. In dramatically changing the structure of communities of large animals, many of which were herbivores (plant-eaters), they thereby substantially altered plant communities and thus modified entire ecosystems. However, it was the agricultural revolution, some 10,000 years ago, that paved the way for the large-scale impacts that threaten the integrity of ecosystems (and thus of the critical services they supply to humanity) today. In the past 150 years, since the Industrial Revolution hit its stride, the scale of the human enterprise (as measured by overall energy use) has multiplied roughly 20 times, as the result of a 4-fold increase in population size and a 5-fold increase in per capita consumption. This, of course, has resulted in a roughly 20-fold impact of *Homo sapiens* on ecosystems.

II. CURRENT IMPACTS

Humanity has become a global geophysical force, altering Earth on a scale as large or larger than forces such as volcanism, erosion, natural oil seepage into oceans, and, especially, the natural extinction of populations and species of nonhuman organisms. Humans have now altered every cubic centimeter of the biosphere, at the very least by changing the climate and distributing synthetic organic chemicals, novel radionuclides, and particulate matter across the globe. This modification of ecosystems has been primarily mediated through its effects on their living elements.

A. Habitat Alteration

The most significant way in which people influence ecosystems is by degrading the habitats required by nonhuman organisms. Often this occurs simultaneously through a variety of channels, so that the types of degradation discussed here often overlap with one another. On rare occasions, however, habitat alteration may have inadvertent direct beneficial effects for people.

1. Deforestation

Forest destruction is probably the most widely recognized form of ecosystem degradation, since it so dramatically transforms the structure of the habitat. In the past few centuries about one-half of all tropical moist forests have disappeared, and in southeast Asia much of the lowland moist forest is already gone. Large areas of temperate forests have been cleared, but in some areas (such as the United States) there has been substantial regrowth. Many ecosystem services are degraded by deforestation, including flood control, maintaining the appropriate gaseous content of the atmosphere (especially through carbon sequestration), regulating the climate, control of vectors of disease, and provision of timber. The reduction of populations of dangerous predators such as tigers is a rare and debatable benefit to *Homo sapiens*.

2. Forest Fragmentation

The increasing patchiness of forests is receiving growing attention, since brute rates of deforestation underrate the ecosystem impact of deforestation for a number of reasons. First, fragmentation leads to a loss of species diversity due to species–area effects. The number of species that a parcel of habitat can maintain is generally a function of its area. In general, 10 fragments averaging 1 km^2 each will retain less diversity than a single fragment of 10 km^2. Second, fragmentation increases the amount of forest edge relative to forest interior, making forest interior organisms more vulnerable to edge-specializing predators. For example, there is some evidence that predation on the eggs of ground-nesting temperate forest birds increases with closeness to edge. Third, fragmentation, especially of tropical moist forests, may lead to further forest destruction because of the penetration of winds and drying into much of the interior of small patches. Fourth, forest fragmentation can lead to secondary extinctions resulting from the loss of biotic interactions. For instance, the extermination of populations of required pollinators can lead to the loss of plants species still present in the fragments. And fifth, fragmentation often results in the loss of species from the fragments due to inbreeding depression (loss of genetic variability leading to lowered fertility or other deleterious effects) or chance (stochastic) extinctions stemming from the small size of residual populations.

3. Selective Harvesting of Forests

Highgrading is the removal of commercially valuable species while bypassing others, and this practice can dramatically alter the taxonomic composition and structure of forests. Removal of mahogany trees is a classic example; so is removal of key understory elements, such as the rattans now being extracted from the forests of Southeast Asia.

4. Fire Suppression

Intervention to prevent fires in forest and chaparral ecosystems that normally burn periodically leads to a buildup of fuels that eventually can result in uncontrollable and extremely hot fires that change the entire system by killing organisms that would ordinarily survive the smaller, cooller fires characteristic of the original fire regime.

5. Conversion to Farms

Clearing of natural ecosystems to introduce crop agriculture normally results in the near total destruction of the ecosystems. Virtually all of the larger native plant and terrestrial animal species are directly extirpated, and many of the smaller plants and animals (e.g., insects, reptiles, amphibians, birds, and bats) go extinct because of loss of habitat and changes in the microclimate.

6. Conversion of Grazing

Conversion of natural ecosystems to grazing can be as or more destructive as conversion to crop agriculture if, for example, large areas of tropical moist forest are cleared to make pastures. On the other hand, if a savanna ecosystem with abundant natural grazers simply has the composition of the grazing community changed (as when small numbers of cattle are added to the ungulate population of an East African savanna), the damage can be much less. Nonetheless, at a minimum the composition and structure of the flora are altered by the introduced herbivores, and often attempts are made to control potential predators on the grazing animals, which may also have cascading effects on the original flora and fauna.

7. Conversion to Infrastructure

The building of structures, highways, malls, and so forth competes with farming for having the greatest impacts per unit area on natural ecosystems. At best, only disconnected fragments of the original system remain, and the delivery of ecosystem services is almost totally disrupted. Conversion to infrastructure, with its destruction or covering of the soils that are normally essential to maintaining a terrestrial ecosystem, is one of the most irreversible forms of human impact on ecosystems. Division of habitat by roads is also a form of conversion to infrastructure. Roads cause impacts on ecosystems far in excess of those created by the road building itself. First, they may form barriers to the normal movement of animals. Second, they serve as conduits for the introduction of invasive organisms (including human beings). Third, dirt roads may influence the flora and fauna of adjacent areas by deposition of dust (which, for example, may function as an insecticide). And fourth, especially in tropical moist forest areas, roads increase penetration, settlement, and recreational use by human beings, all of which generate additional impacts.

8. Mining

Extraction of minerals varies in its ecosystem impacts, depending on the scale of the mines, whether they are underground, open pit, or placer (hydraulic/washing/dredging) operations, and the material mined. Large-scale open-pit mining tends to be totally destructive and expensive and very difficult to ameliorate. Underground mining usually has less impact, but can destroy aquatic systems over large areas through the runoff of toxic drainage from the mines, and it converts terrestrial systems into "slime pits" for mine wastes.

9. Damming and Other Water Development

Human attempts to intervene in the hydrological cycle have massive impacts on the ecosystems of rivers and streams by changing water flow patterns and temperatures and by blocking the movements of anadromous fishes (ones that breed in streams but spend much of their lives at sea). Conditions may be created that encourage the spread of disease, as in the increase of *Bilharzia* in Egypt after the building of the Aswan High Dam. This dam made possible the construction of irrigation canals that became ideal breeding grounds for the parasites that cause the disease. Terrestrial ecosystems, of course, are directly destroyed by inundation.

10. Wetland Drainage

The drying of wetlands eliminates the organisms and services supplied by estuaries, marshes, and bogs (flood control, water purification, and nursery services for commercially important fish species). Loss of wetland habitats also disrupts the migratory patterns of birds and reduces their food supply, including wildfowl populations valued by hunters.

11. Siltation of Onshore Ocean Waters

A flow of silt into the oceans results from timbering operations, construction, and other human activities that occur in coastal zones. In the tropics, siltation leads to the destruction of coral reef ecosystems by the smothering of coral organisms. This in turn leads to the loss of local fisheries and ecotourism opportunities.

12. Toxins in Aquatic Systems

The toxification of onshore ocean waters, lakes, and streams may disrupt marine ecosystems and negatively impact fish populations (by direct poisoning, or alteration of food chains). Toxins may also render fish unfit for human consumption, thus disrupting the food supply service of aquatic ecosystems. Sewage dumping has converted several substantial areas of the oceans into "dead zones" in which oxygen levels are so low that most animals cannot live in them. Toxification, which has influenced the entire planet from pole to pole, is one of the most subtle forms of habitat alteration. Often toxic substances are active in concentrations that are difficult to detect, especially when they function by mimicking hormones that influence the early development of animals. The long-term ecosystemic effects of adding tens of thousands of synthetic organic chemicals to the global environment are simply unknown.

13. Ozone Depletion

Thinning of the ozone layer adds another toxic substance to Earth's ecosystems—an increased flux of dangerous ultraviolet B radiation (UV-B). The impact of such radiation can be judged from the fact that until some 400 million years ago life could not exist on the land surface because there was no ozone layer to screen out the UV-B. Until then, organisms had sheltered in aquatic ecosystems, because water rapidly reduces the UV-B flux. Ozone depletion has been most serious in the region of the "ozone hole" in the Antarctic, where impacts on oceanic ecosystems are already reported. As with other toxic substances, the long-term effects of moderate ozone depletion are difficult to predict. It has been postulated that increased UV-B or other toxic materials are involved in a global decline of amphibian populations, but the data are still too fragmentary to permit definitive judgments to be made.

14. Recreation

Recreational activities have diverse impacts on ecosystems. Off-road vehicles compact and erode soils and kill organisms that burrow in them, and are especially destructive in arid climates. Trails in parks can cause local erosion and disrupt the normal activities of wildlife. Boat anchors and scuba-diving activities contribute substantially to the destruction of coral reefs in some localities. Ecotourism, which can provide economic benefits for conservation, can also be destructive if not properly controlled. The outflow of sewage from coastal hotels pollutes the waters of offshore reefs. The disturbance of animal communities and native flora by tourist vehicles in some African game parks is another example of recreational impact.

15. Climate Change

Alteration of the climate, which may be induced or accelerated by human activities, is a major mechanism for altering ecosystems. Global warming appears to be already changing both terrestrial and aquatic ecosystems in the boreal regions. Perhaps the most serious threat is an apparent thinning by one-third of the sea-ice in the Arctic. Shrinkage of the ice pack will have large, if difficult to predict, consequences for polar bears and other aquatic mammals and the food chains in which they participate. But the greatest threat of broad-scale ecosystem impacts doubtless resides in the potential for changing the patterns of oceanic currents. For example, should the Gulf Stream off the U.S. Atlantic Coast fail or be deflected, all of the ecosystems of western Europe would be severely stressed, to say nothing of the human populations there.

Anthropogenic climate change could be many times faster than the relatively rapid natural changes that occurred following the last glaciation. It would certainly cause large-scale population extinctions in some areas, as already appears to have been detected in butterfly species in western North America and Europe. It might also lead to a substantial loss of species diversity. Much will depend on the speed and nature of the changes, and on the capability of organisms to migrate past extensive barriers created by near ubiquitous human development.

B. Alien Invasions

Alien invasions are little appreciated as engines of ecosystem modification, and yet they are one of the most important.

1. Islands

Many oceanic islands have lost most of their native floras and faunas to species transported by *Homo sapiens*. The destructive invaders include goats (which long ago played a major role in transforming the ecosystems of the Mediterranean basin), rats, mongooses, Philippine brown snakes (which devoured the avifauna of Guam), and mosquitoes (which carried bird malaria to Hawaii, resulting in the extinction of almost all species of native birds).

2. Continents

Continental ecosystems have also been highly modified by invasions. Most of the grasslands and woodland savannas of California are now nearly devoid of native herbaceous plants, having been substantially taken over by introduced annual grasses and herbaceous weeds from Europe.

3. Onshore Waters

Not only has the flora of California been greatly modified by exotics, but the coastal waters, especially around San Francisco Bay, are host to a myriad of invaders. Many of these arrived in the ballast water of ships arriving from all over the world. One, the eastern shipworm, was tolerant of the relatively dilute seawater of the Bay, as the native western shipworm was not. After it was introduced early in the twentieth century, it proceeded to destroy the wooden piers and bridges that laced the Bay, doing as much financial damage as the 1906 earthquake.

4. Disease Organisms

Pathogens moved about by human beings also greatly influence ecosystems. A classic case was the introduction of a virus disease of ruminant animals into northeast Africa around 1884. The disease raced south, reaching South Africa in 1896. It not only decimated the cattle herds of peoples like the Masai (causing severe famines), but also infected giraffes, buffalo, and wildebeest, causing a die-off that led to starvation of their natural predators. One consequence was that lions switched to eating people, so that farmers deserted their land. In the absence of farmers and native browsing animals, brush and woodlands invaded grasslands, altering the entire ecosystem. Some resistance developed in the animals, and the situation has fluctuated ever since, with rinderpest remaining a factor in East African ecosystems.

C. Overexploitation of Ecosystem Components

Habitat alteration and invasions are not the only ways in which human activities produce impacts on ecosystems. Another important pathway is the overexploitation of economically desirable organisms, which in turn alters habitats.

1. Food

Often overexploitation occurs when natural populations are harvested for food. A classic example was the persecution of the passenger pigeon, which once was the most abundant bird in North America. One flock was estimated to contain two billion birds. Flocks moved around, breeding when large crops of beechnuts, acorns, and other forest nuts were found. They were commercially hunted to extinction to supply food markets in eastern cities; the last individual died in a zoo in 1914. It has been suggested that their absence altered the eastern forest ecosystems and greatly increased the food supply of rodents, which are now implicated in the cycle of lyme disease. If true, the ecosystemic effect of their extermination had an impact on humanity beyond the loss of a once abundant and easily harvested source of animal protein.

Fishing often alters ecosystems by disrupting normal oceanic food chains, especially when there is overfishing and when there is heavy "bycatch" of nontargeted animals (some 27 million metric tons of marine life are killed and discarded by oceanic fishing fleets, equivalent to about a quarter of the total catch retained, and this does not account for the heavy local mortality caused by dynamite and cyanide fishing on coral reefs). Increasingly, trawls dragged along the bottom are destroying the biotic structure there, with unknown future consequences.

2. Natural Products

Demand for natural products can also lead to overexploitation with accompanying ecosystem alteration, as discussed earlier under forest clearing and fragmentation. But it is often difficult to measure the impact, for example, of the harvesting of specialty woods by "high-grading." Again, far more damage to the forest is sustained by the building of roads, the associated felling of nontarget trees, and the transport of harvested trunks by "skidding" compared to the small amount of damage from the selective removal of valuable hardwoods. Similarly, although the demand for sea turtle shell and soup has had serious consequences for turtle populations, we know little about the ways oceanic ecosystems may be altered by the reduction of turtle populations.

3. Esthetic Resources

The demand for pets frequently leads to the overexploitation of natural populations. Some of the cyanide fishing on coral reefs is aimed at aquarium fishes rather than fishes sold to grace the plates of rich Hong Kong businessmen. Freshwater tropical fishes are also often overexploited for the aquarium trade. The ecosystemic changes being caused by the latter have not been seriously investigated. The pet bird trade has large-scale impacts on some species and in some regions. This trade is so popular in Indonesia that songbirds are virtually absent from much of Java because of trapping (and overuse of pesticides). In the Pamulka Passarum bird market in Jakarta, some 60,000 individuals and 160 species are for sale on a given day. But we do not even know the impact of bird removal on the insect populations of Java, let alone the overall (or long-term) ecosystemic consequences. The story is similar for the removal of giant saguaro cacti in Arizona or the poaching of orchids, chameleons, and geckos around the world—biologists realize the ecosystems are being altered, but there is neither the time nor the personnel required to evaluate these and many other impacts.

4. Scientific and Educational Use

Finally, sad to say, some organisms are overexploited for scientific and educational purposes. A classic example is the grass frog, *Rana pipiens*, which once was abundant and harvested in large numbers for use in dissections in high school and college biology courses. It is now a rather scarce organism.

III. SIGNS OF ECOSYSTEM DETERIORATION

If humanity is causing substantial impacts on ecosystems, can we detect what they are? In many cases, the answer is yes, although precise measurements are difficult to obtain, mainly because baseline measurements were never taken before impacts occurred. Most, if not all, of the observed and measurable impacts are negative; positive changes have been observed largely in cases where the pressure has been lightened (e.g., the recovery of forests in parts of the eastern United States after failing farms were abandoned).

Perhaps the most direct way to measure ecosystem deterioration is to measure an area's biological productivity (net primary productivity, NPP), if possible over time. If an area is converted from a natural to a human-directed system (forest to pasture; meadow to farm fields), NPP usually falls. A further decline in NPP (for instance, crop yields) over time indicates deterioration. Such declines, even in the face of increased fertilizer applications, have been widely observed, and it is well known that fertilizer use can mask losses of micronutrients from soil and declines in soil texture for considerable periods before declining productivity becomes apparent. In such situations, not only is crop or forest productivity lost, but the ecosystem services of soil replenishment and nutrient/waste recycling (which are also a vital part of the planet's biogeochemical cycles that control the distribution of nitrogen, oxygen, carbon, and water, among others) are usually damaged.

A. Faltering Food Production

The decline of food production may be the most significant trend in which human impacts on ecosystems play a role. Following unprecedented rates of growth in food production after the middle of the century, the global grain harvest has failed to increase on a per capita basis since 1984. Since cereal grains represent the human feeding base and are equal in weight to roughly half of all foodstuffs produced by agriculture, trends in grain harvests are the best indicator of food supplies and availability, although economic factors and changes in eating habits are also significant. About a third of the world grain harvest is used for feeding livestock, and shortages in grain supplies for human food are often compensated by reducing its use as feed. Grain shortages often cause rises in price on the global grain market, which in turn cause the prices of meat and other animal products to rise and consumer demand to fall. Similarly, reduced demand for animal products, as has occurred in some industrial nations in recent years, leads to reduced demand for feedgrains. In the United States, demand for animal products has shifted away from beef to poultry, which can be grown with about a third as much grain per pound of meat produced. In the industrial countries, more than half the grain consumed is used for feed. In developing nations, by contrast, most grain is directly consumed by people. But in some countries (especially China) this is changing, and demand for animal foods is rising rapidly.

Despite these shifting economic and consumption patterns, the overall trend of food production has been increasingly problematic in recent years, with aggregate production failing to keep abreast of population growth, which itself has slowed markedly in the past decade. Production in sub-Saharan Africa has fallen especially far behind. The reasons are many and vary from area to area. They include the Green Revolution running out of steam, diminishing returns from fertilizer applications, decline in crop genetic diversity, loss of prime farmland to urbanization, and land degradation from poorly managed irrigation. Among those of rising importance are underlying trends of land degradation and the related loss of ecosystem services. It may well be that potential food production has actually kept up with population growth. But key issues are: How much higher would that potential be if important ecosystem services had been carefully conserved? And can humanity depend on linear extrapolations of recent production trends over either a 50-year or decadal timescale?

Of the ecosystem services that are being impaired or lost, some are basically local—soil erosion and depletion, loss of natural pest and crop disease controls, and loss of pollination services are obvious examples. Others are regional, such as the consequences of damage to nearby watersheds. Thus deforestation of a watershed area leaves adjacent or downstream areas more vulnerable to floods and droughts. The destruction by Hurricane Mitch of a substantial portion of the agricultural capacity of Central America in the fall of 1998 is a case in point. There also is evidence that maintaining natural areas interspersed among intensively farmed areas helps to preserve soil fertility as well as provide sources of natural pest control, pollinators, and impediments to the spread of diseases.

The Green Revolution, along with agricultural expansion, a dramatic increase in irrigated lands, and other factors, allowed grain production to rise by more than 73% from 1960 to 1980, while the population expanded by 46%. But from 1980 to 1997, grain production rose only 31%, just equal to the population's growth of 31%. Up to 1984, grain production stayed ahead of population growth, reaching a per capita peak then of 342 kg; since then it has fluctuated between 301 and 335 kg per capita.

The degree to which human impacts on ecosystem services are involved in the difficulties of agriculture is difficult to evaluate. Some crop damage by pests is clearly due to the diminution of natural pest control services, but pesticide use and cultural practices make up for some of the loss. Similarly, declines in natural pollinator services are occurring, but data are lacking on the degree to which exotic pollinators (honeybees in many areas) are able to take up the slack. Loss of flood control services, due primarily to deforestation of watersheds, often results in serious damage to crops in places as diverse as the Indian subcontinent and Central America. Soil loss to erosion—perhaps 25 billion tons annually in excess of soil regeneration—also reduces agricultural production. And the anthropogenic modification of ecosystems is depleting the genetic library of crop relatives, crop pollinators, and crop biocontrol agents, which are all critical to maintaining crops in their coevolutionary races with diseases and pests.

It is important to remember that agriculture is largely dependent on the weather, and that climate stabilization is a vital service of ecosytems. Thus, for example, severe damage to agriculture in Central America from Hurricane Mitch may have been exacerbated by a decline in this service if anthropogenic global warming played a role in the unusual behavior of the storm. Floods and droughts, and changes in the timing of the onset of seasonal warming trends or rains, will obviously have major impacts on agriculture around the world.

B. Fisheries Decline

Fisheries yields tell an even more dismal story—and at sea there is no doubt that yields have not been keeping pace with demographic expansion. Some two-thirds of the world's major fisheries are being maximally harvested today or are in decline. A per capita peak in fisheries yield was reached in 1988 at 17.2 kg per capita, but yields have failed to meet that level for the past decade. Most of the blame for declining yields can be placed on overharvesting, yet more systemic environmental damage has also played a role through the pollution, modification, and destruction of estuaries, coral reefs, mangrove fringes, and coastal wetlands. Anadromous fish (such as salmon) have suffered from the damming of rivers and the siltation from bank erosion. Major oil spills have taken a toll on seabirds, marine mammals, and shellfish, causing lingering damage to coastal fauna in many areas. Aquaculture (fish farming) harvests, while offsetting declines in traditional fishery yields, have caused new pollution and habitat destruction problems that affect natural fish populations, and they rely extensively on fish products for feed. Thus aquaculture tends to produce a net loss in fish biomass. It also creates a rising demand for feedgrains and other agricultural products to support their yields, thus competing, along with livestock operations, with human beings for the food they raise.

C. Change in the Epidemiological Environment

Disruptions in the epidemiological status quo may be another sign of ecosystem modification. Anthropogenic ecosystem changes are clearly threatening the deterioration of the climate stabilization service. Those changes may have played a role in the warming that has allowed the mosquitoes that transmit dengue fever to move northward in North America. This warming is expected to have similar effects on the wider occurrence of malaria and other tropical diseases.

D. Decline in Water Quality

Increasing water pollution is often a direct reflection of the modification of ecosystems. A classic example is that of the New York City water supply. Around 1900, New York City's water was of such high quality that its was bottled and sold throughout New England. Recently the U.S. Environmental Protection Agency has notified the city that its water had dropped below acceptable quality. The reason for this decline was the degradation of the Catskill Mountains watershed that served the city—fertilizer runoff and inadequate local sewage treatment had impaired the natural water-supply service. It was estimated that constructing a plant to treat the Catskill water would have a capital cost of $6–8 billion, with roughly a $300 million annual maintenance cost. In contrast, restoring the functioning of the natural Catskill ecosystem would cost only $1–1.5 billion. This is a case of ecosystem modification and the loss of essential ecosystem services.

E. Other Symptoms

There are, of course, many other readily observable symptoms of ecosystem deterioration caused by human action, some of which have already been mentioned. On land they include flooding, landslides and heavy soil erosion, changes in microclimate, the local disappearance of sensitive species of plants and animals, siltation of streams, and lowered water tables. A general decline of amphibians appears to be occurring over much of the world, and may be caused by multiple disruptions from many causes. In oceanic systems, changes in the mix and abundance of animal, plant (e.g., kelp and seagrass), and planktonic species are also observable and often, as in cases of algal blooms or precipitously declining fisheries yields, are signals of serious disruption.

IV. DRIVING FORCES

The impacts of human beings on ecosystems are a product of three multiplicative factors: population size, affluence (or per capita consumption), and the technologies and social-political-economic relations that supply the consumption. These relationships are often condensed for convenience into the "IPAT equation"

$$\text{Impact} = \text{Population} \times \text{Affluence} \times \text{Technology} = I = P \times A \times T$$

If one uses per capita energy consumption as a surrogate for A × T, the result is that quoted in the beginning of this article—a 20-fold increase in the human impact on ecosystems over the last 150 years.

A. Overpopulation and Continued Population Growth

Population growth is probably the single most important factor leading to impacts on natural ecosystems. Early in the last century there were only 1 billion people. That number had doubled to 2 billion around 1930, doubled again to 4 billion in 1975, and reached 6 billion in 1999. In other words, it took on the order of 300,000 years for the population of *Homo sapiens* to reach a billion, a single century to add the second billion, 30 years to add the third billion, 15 years for the fourth, 12 years for the fifth, and another 12 years for the sixth billion. Simply adding so many people to the population required gigantic interventions in natural ecosystems in order to bring sufficient land under cultivation to feed them. This effect was exacerbated because each individual added to the population tended to have a disproportionate ecosystemic impact. Naturally people farmed the most fertile soils first, got their water from the nearest sources, mined the most concentrated ores, and so on. More people required the farming of more marginal land, and thus more land per person. The transportation of water over greater distances involved using more water per person to make up for losses in storage and transport, and the disruption of more area per person for dams and other infrastructure. The mining of poorer ores means that more ecosystemic destruction by mining was caused per person, since more ore must be dug up to provide the same level of metal use for each additional individual.

Since at least 1950 the entire world clearly has been overpopulated by the simple standard that the human population could not be sustained by the flow of resources generated by incoming sunlight mediated through farms, forests, the hydrological cycle, and so forth—the agricultural and natural ecosystems of Earth. Instead of preserving humanity's "natural capital," it has been necessary to expend it. Three elements of that capital are especially important. The first is deep, rich agricultural soils, which are generated on a timescale of centimeters per millennium and are now in many areas being destroyed at a rate of centimeters per decade. The second is aquifers, many of which were last filled during the ice ages and which are now, in many if not most areas of the world, being pumped at many times the natural recharge rate. The third element is biodiversity, the species of microorganisms, plants, and animals that are working parts of ecosystems, which is now being exterminated at a rate that is unprecedented in the past 65,000 years—at perhaps 1000 times its natural regeneration from speciation.

Furthermore, the high *rate* of addition to the human population in the second half of the twentieth century meant that little concern could be shown for the possible long-term effects of the interventions designed to keep people eating. Programs such as the "Green Revolution" (the exportation of high-yield agriculture from rich temperate zone nations to poor tropical ones) were driven in part by the fear of massive starvation, and their consequences for natural ecosystems and the vital services they provide were largely ignored. So were their social impacts.

B. Overconsumption

Overconsumption (consumption far in excess of basic needs) is almost as important as overpopulation as a driver of ecosystem modification. For example, a significant proportion of the ecosystem modification caused by the agricultural enterprise is created by growing grain and feeding it to animals destined for human consumption. This is the case even though in rich societies the consumption of products containing animal fat (and, perhaps, overconsumption of animal protein) causes serious public health problems. The directly toxic and hormone-mimicking effluents from the production of myriad products, many of them plastic or packaged in plastic, also are doubtless having ecosystemic effects, although data beyond the effects on single species are lacking. But perhaps the most serious overconsumption problem derives from converting societies to dependence on the automobile, combined with the marketing and purchase of cars that are much too large, heavy, and inefficient for the purposes to which they are put.

C. Use of Environmentally Malign Technologies

The use of technologies that are unnecessarily environmentally malign is related to overconsumption. The use of inefficient automobiles is mirrored in the use of unnecessarily inefficient lighting, heating, appliances, and so on. But the basic technological problem in society at the end of the twentieth century is the persistent overdependence on fossil fuels, and especially coal, as an energy source.

D. Faulty Economic Arrangements

Market failures are partly to blame for the overdependence on fossil fuels, because the market prices of those fuels do not come close to reflecting their social costs. For example, the costs incurred by society in the form of lung disease from the effluents created by fossil fuel burning are not factored into the market price. More importantly, neither are the present and potential costs to society of climate change induced by carbon dioxide emissions from fossil fuel use. If these and other social costs (such as the pollution due to oil spills and the discarding of lubricants used in automobile maintenance, other emissions, and medical costs deriving from automobile accidents) were captured in market prices, the price of gasoline in the United States might be quadrupled. This could have the beneficial result that fuel-efficient cars would be produced and then more people could use them, in the process creating less pollution and reducing the demand for environmentally destructive oil extraction.

E. Faulty Social Arrangements

Social arrangements often work against the protection of ecosystems. For instance, society has not designed educational systems to apprise its citizens of the biophysical and social realities that it faces. Few people, even "well-educated" people in rich societies, can provide a coherent explanation of ecosystem services or the threats to their delivery. Indeed there is a general failure of the media and standard educational system to make explicit the connections among the many factors and effects of human alteration of ecosystems.

F. Faulty Political Arrangements

Political biases and errors also play a substantial role in endangering ecosystems. Societies generally lack "foresight" institutions that have the capacity and independence to analyze complex problems and provide competent and independent advice to government and the general public. In addition, there is the age-old problem of disparities in power that now have increasingly widespread—often global—implications for the sustainability of society. Powerful supporters of politicians in Rome, before its collapse, were hard-pressed to weaken the Empire. Modern business interests who press the politicians on their payrolls to take no actions to ameliorate global warming could theoretically bring down our current civilization.

V. SOLUTIONS

In light of the myriad uncertainties about both human and ecosystem behavior, it would seem prudent to try to reduce the three factors in the IPAT equation simultaneously, rather than risk a possible catastrophic collapse of services as *Homo sapiens* progressively alters natural ecosystems. What sorts of steps might be taken to avoid such a collapse? Some are proximate—steps that would directly address impacts on ecosystems. Others are ultimate—steps that would deal with the basic drivers of ecosystem deterioration.

A. Proximate Solutions

A nonexhaustive list of proximate steps might include:

1. Limiting development so that a minimum of the remaining relatively undisturbed ecosystems would be destroyed. In virtually every country where new infrastructure is needed, it should be provided as far as possible by redeveloping intensively areas that are already highly disturbed.
2. Agreements to slow climate change need to be solidified rapidly, despite the grave political and economic difficulties of doing so. The international protocol to maintain the ozone layer should be reexamined in light of recent data to see that its performance is adequate.
3. Establishment and protection of reserves designed to protect relatively undisturbed ecosystems and the elements of biodiversity they contain. This should only be done when it is possible to integrate the interests of local people into the reserve design—otherwise in the long term the effort is likely to be wasted as population growth pushes people increasingly into reserves that they do not value as such.
4. Limit toxic releases by shifting the burden of proof to those who would claim that the introduction of a novel compound into use (and thus into the environment) carries benefits that clearly exceed the social costs.
5. Sustainable-yield harvesting should be enforced for all living resources. Harvesting systems should be designed with substantial attention to the precautionary principle (erring on the side of conservatism) in order to buffer society against the consequences of overoptimistic estimates of what is sustainable.
6. Apply countryside biogeography principles. Countryside biogeography is the nascent field that is attempting to develop principles for making already disturbed areas most hospitable for biodiversity and thus more likely to provide necessary ecosystem services. One step in doing this would be to pass (or strengthen) laws to protect endangered populations, species, and landscapes, and to restore moderately degraded areas, including restoration with useful, native species.

B. Ultimate Solutions

Steps to deal with the ultimate drivers of ecosystem deterioration must address the factors of the IPAT equation.

1. Population Reduction

Population growth should be halted as soon as humanely possible, and then a slow decline can be initiated toward a level that can be sustained indefinitely at whatever average level of consumption is selected. Selection of that level, which is closely related to the issue of optimum population size, need not be debated in the near future. Population size will change with available technologies and preferences. In the foreseeable future the human population will be living above the world's long-term carrying capacity, so that a goal of simply halting growth is likely to be valid for at least a half century. After this period, there will be abundant time for scientific study and public debate over where a decline should be halted. Human beings, as far as can be told, have always intervened to adjust population sizes to needs perceived primarily through the situations of reproducing couples. In recent decades there has been a shift to substantial consideration of the needs of society as well, and that trend should be accelerated in the future.

2. Consumption Control

Consumption control will be a more difficult task than controlling population growth, if current trends and attitudes are any guide. This is an area fraught with difficult issues of equity and feasibility. For instance, if the gasoline tax in the United States were raised to the point where gasoline was as expensive as in Europe, a substantial decline in fuel consumption would be likely. But there would be a deleterious impact on poor people who must commute to work by car. Compensatory tax changes would be needed to lighten that burden.

Not only do many of the things that human beings now consume seem unnecessarily harmful, but the quantities consumed per person's often seem outrageous. But since one person's outrage is triggered by the use of something that another values greatly, it is difficult to imagine how consumption patterns can be changed significantly without substantial social conflict. However, much can be accomplished by working hard to increase the energy and material efficiency of the processes by which consumption is served—such as using bikes, mass transit, and light, fuel-efficient automobiles to "consumers" of travel, the more efficient production of consumer products, the reduction of packaging, and so on.

3. Substitution of Technologies

Furthermore, it is possible to substitute environmentally benign technologies for currently used environmentally malign technologies. For example, with proper education, natural pest control services, enhanced by integrated pest management, can be substituted in agriculture for the broadcast spraying of toxic pesticides. Buildings can be equipped with "more expensive" energy-saving devices, such as video conferencing and telecommunications, that can be increasingly used to eliminate energy-expensive travel to meetings and workplaces and would pay for themselves through energy savings within one to five years.

4. Revising Socioeconomic and Political Systems

All of the foregoing steps would benefit from public education and broad social discussion of the need to, as the economists say, "get the prices right." That is to say, the prices of goods and services need to be adjusted so they more closely reflect the social costs of each. In other words, external costs should be internalized. The heavy burden placed on society in environmental deterioration and public health from, for example, the overuse of large automobiles should be reflected in the prices of both the vehicles and the fuel they consume.

Success at evaluating and implementing the steps that can and should be taken in various nations and in diverse economic strata within nations may depend on substantial changes in governmental institutions and the generation of broad social consensus. Some revision of socioeconomic and political systems is likely to be required if humanity is to preserve the ecosystem services upon which it ultimately depends, and the biodiversity upon which the systems themselves depend.

See Also the Following Articles

AGRICULTURE, INDUSTRIALIZED • DEFORESTATION AND LAND CLEARING • DISEASES, CONSERVATION AND • ENERGY USE (HUMAN) • FISHING, COMMERCIAL • ISLAND BIOGEOGRAPHY • LAND USE ISSUES • PLANT INVASIONS • POLLUTION, OVERVIEW • POPULATION STABILIZATION (HUMAN)

Bibliography

Daily, G. C. (1995). Restoring value to the world's degraded lands. *Science* 269, 350–354.

Daily, G. C., (ed.). (1997). *Nature's Services*. Island Press, Washington, D.C.

Ehrlich, P. R., and A. H. Ehrlich (1991). *Healing the Planet*. Addison–Wesley, Reading, Massachusetts.

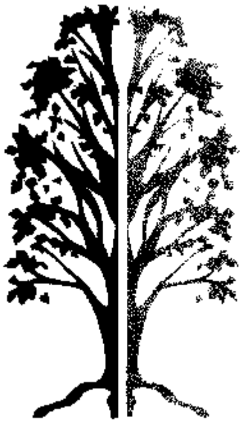
Ehrlich, P. R., and A. H. Ehrlich (1996). *Betrayal of Science and Reason: How Environmental Anti-science Threatens our Future*. Island Press, Washington, D.C.

Postel, S. L., G. C. Daily, and P. R. Ehrlich (1996). Human appropriation of renewable fresh water. *Science* 271, 785–788.

Turner, B. L., II (ed.). (1990). *The Earth as Transformed by Human Action: Global and Regional Changes in the Biosphere over the Past 300 Years*. Cambridge University Press, Cambridge, United Kingdom.

Vitousek, P. M., H. A. Mooney, *et al.* (1997). Human domination of earth's ecosystems. *Science* 277, 494–499.

Wilcove, D. S., D. Rothstein, *et al.* (1998). Quantifying threats to imperiled species in the United States. *BioScience* 48(8), 607–615.



HUMAN IMPACT ON BIODIVERSITY, OVERVIEW

Leslie E. Sponsel
University of Hawaii

GLOSSARY

cultural ecology Analysis of how culture influences the interactions between a human population and the ecosystems in which they reside; also called ecological anthropology.

culture System of socially learned, shared, and patterned ideas, institutions, behaviors, and their material products that distinguishes a particular society.

diversity principle General geographical coincidence between high concentrations of both biological and cultural diversity, usually in the tropics.

ecological transition Tendency for societies to be in growing disequilibrium with their biophysical environment as they increasingly deplete natural resources and degrade their habitat, thereby reaching new thresholds of environmental impact.

historical ecology Transdisciplinary and diachronic analysis of how human societies and ecosystems change and in turn transform one another through time in local and regional landscapes. Data are drawn from geology, archaeology, history, and other sources.

swiddening Umbrella term including diverse types of horticulture in which a small section of forest is cut and burned to plant crops in a temporary garden; shifting cultivation is used as a synonym, but slash-and-burn cultivation is now considered to be a pejorative term.

traditional or local environmental knowledge (TEK) Detailed and accurate knowledge about the environment, including biotic species and ecological processes, that many indigenous and other peoples have developed, accumulated, and apply in their daily and intimate interactions with their habitats and in their system of natural resource use and management. The study of TEK is included within ethnoecology.

THROUGHOUT HUMAN PREHISTORY AND HISTORY, HUMAN IMPACTS ON BIODIVERSITY have reached progressively higher thresholds. Most likely the net impact of humanity has been to reduce biodiversity. However, at the population level, the types and magnitudes of human impacts on biodiversity vary tremendously through time and space, depending on the specifics of the particular context. Many societies have decreased local biodiversity, whereas many have sustained or even increased it. Nevertheless, because of the widespread direct and indirect impacts of humans on biodiversity, any consideration of biodiversity must also assess the possibilities of human influence.

I. PRINCIPLES

A. Human and Environmental Diversity

Not all humans are equal in their impacts on biodiversity. While it is widely recognized that there is tremendous diversity in nature, there is also considerable diversity in humanity. Nearly 7000 distinctive cultures exist today and there were many times more in the past, probably on the order of at least tens of thousands over some four million years of human existence. Accordingly, there is tremendous diversity in human relationships with and impacts on biodiversity. In examining human impacts, it is imperative to consider the diversity of humans through time (prehistory through history) and space (cultures and regions).

Environmental diversity is another variable in determining the human impact on biodiversity. Some environments are simply much more fragile and vulnerable than others, especially those that are relatively simpler, such as the arctic and deserts, and those that are isolated with a high proportion of endemic species, such as oceanic islands. Also, some regions are simply much less accessible or hospitable to humans than others, including the poles, mountain heights, deserts, and deep oceans. In these cases, the label "wilderness" may still apply—pristine nature that is largely unaffected by humans.

Furthermore, given the dynamism of ecosystems and ecological processes, together with the widespread impact of humans on them, it is misleading to consider most environments as pristine, virgin, primeval, or wilderness. The condition of any ecosystem is the cumulative product of previous conditions, usually including human impacts. The more important question is not whether or not a human society had or has any environmental impact, but its particular forms and magnitude, and if negative then the extent to which it is reversible and allows natural regeneration within a normal period of time.

In studying human impact it cannot be assumed that the previous environment was in a state of equilibrium or climax stage in ecological succession. Furthermore, even when humans are a factor in environmental change, it cannot be assumed that other agencies, such as natural alterations in climate or weather and natural fires, are not also contributing to change.

In any case, *it is unrealistic to consider biodiversity anywhere without also considering the possibilities of human influence.* Human impact is inevitable because, ultimately, biodiversity is the primary or raw natural resource that all societies rely on for their subsistence and economy. Also, biodiversity is the biotic component of ecosystems and ecological processes on which human survival, adaptation, and welfare ultimately depend.

B. Types of Impact

Human impact on biodiversity, direct or indirect, involves four basic factors: (1) over-exploitation of natural resources; (2) habitat modification, conversion, and fragmentation; (3) the introduction of exotic (nonnative) species; and (4) pollution. Any one of these four factors may influence ecosystem composition, structure, and function; ecological processes; and biodiversity, especially through species extinction. Documented extinctions during the last 400 years include 484 animal species and 654 plant species, but this must be only a fraction of the actual total (Heywood, 1995: 198). However, in some circumstances the second and third factors may increase biodiversity locally, because some components of biodiversity are anthropogenic (human-caused). In Britain, for example, there has been a net increase in mammalian diversity as 21 of its 49 species are introductions (Heywood, 1995: 757).

Human populations may influence biodiversity at the genetic, population, species, ecosystem, biome, and/or biosphere levels of the ecological spectrum. Because species as well as ecosystems are interrelated and interdependent, human impact on any one species or ecosystem may influence others as well, much like a chain reaction.

C. Species and Population Levels

Globally humankind has increasingly tended toward greater levels of ecological disequilibrium, that is, with—many (but not all) populations exceeding the carrying capacity of their habitat, depleting resources, degrading environments, and thereby threatening and reducing biodiversity. There have been at least seven successive thresholds of environmental impact in cultural evolution, their inception approximated in years before present (B.P.) and varying regionally: the use of fire by foragers (hunter-gatherers) (0.5 million), farming (5000–10,000), cities (5000–6000), exponential growth in the human population (1000), European colonialism (500), industrialization (150–200), and globalization (50). These thresholds are part of a continuum or gradual process, although an accelerating one. (These thresholds are somewhat analogous to the states of water as ice, liquid, and gas. The substance—H_2O—is the same, but a gradual quantitative increase in temperature leads to major qualitative changes in its physical properties.) In the case of socio-cultural evolution, scale of complexity is a pivotal factor (Bodley, 2000) (see Table I).

At the population level, the situation is complicated. Long ago some societies developed relatively sustainable economies that did not markedly endanger or reduce biodiversity, as evidenced by their longevity in prehistoric and/or historic records. Some societies even increased biodiversity. Furthermore, the same society may increase biodiversity in certain ways, yet decrease it in others. Thus, human impact varies in time and space as well as with different species, ecosystems, and biomes. Consequently, any generalizations about human impact must be carefully scrutinized with attention to temporal and geographic scales as well as the specific details of the particular cultural and ecological contexts.

It follows that a holistic anthropological approach, one emphasizing cultural and historical ecology, is indispensable for assessing the possibilities of human impact on biodiversity in any context. It could even be argued that cultural and historical ecology are indispensable for understanding biodiversity itself, because it is so rarely free from some degree of human influence, at least indirect. Indeed, biodiversity as a concept is a Western cultural construct developed at a particular stage in history, even if the associated physical phenomena are undeniably real (e.g., Soulé and Lease, 1994).

TABLE I

General Trends in Sociocultural Evolution Correlated with Increasing Thresholds of Human Impact on Biodiversity

1. Population—nomadic to sedentary settlement pattern with increasing population density, nucleation (settlements to cities), and pollution.
2. Food—wild to domesticated foods with shift from foraging to farming
3. Energy—somatic (human and animal) to extrasomatic (water, wind, wood, fossil fuel, nuclear) sources of energy for work
4. Land—extensive (horticultural) to intensive (agricultural) land use; land tenure—community/public to private/corporate ownership
5. Economy—subsistence (satisfying basic physiological needs) to market of surplus production to materialist consumerism; local self-sufficient to regionally and then globally interdependent economy (globalization)
6. Waste—organic products that are readily biodegradable to more recently those like metals and plastics that disintegrate very slowly
7. Scale—small and decentralized to large and centralized societies (states), the latter with increasing import of natural resources from ecosystems in distant regions
8. Differentiation—egalitarian to hierarchical (stratified) societies, the latter with increasing inequality in access to resources, goods, and services and institutionalized warfare
9. Alienation—daily to occasional contact with and feedback (monitoring human impact) from the natural environment; eventually with alienation from nature and other humans, and increasingly decisions made by agents far removed from the locations they affect
10. Worldview—ecocentric to anthropocentric and egocentric worldviews, attitudes, and values; also sacred/moral to secular/amoral and utilitarian orientation to nature; may include shift from biophilia (love of nature) to biophobia (fear of nature)
11. Balance—some degree of dynamic ecological equilibrium with recognition of limits to increasing disequilibrium with assumption that there are no limits (i.e., ecological transition)
12. Impact—environmental modification to conversion (natural to cultural landscapes) and fragmentation (remnant patches of nature); also toxification with industrialization; local to global impact on biodiversity and environments

D. Ambivalence

Biologists and others have been ambivalent about the human species. Most would readily consider *Homo sapiens* to be a part of nature as a product of biological evolution, but apart from nature in terms of ecology. However, in much of the past and in many cases to this day, human populations are simply one component in the dynamics of ecosystems and ecological processes. The idea that humans are unique and apart from nature may derive from the anthropocentrism (human centeredness) of Western civilization and of Christianity, Judaism, and Islam. (Anthropocentrism is also a contributing cause of the environmental crisis.) Actually, every species is unique as a closed genetic system and in other respects. However, it may be valid and useful to view *H. sapiens* as the dominant or most important keystone species in many if not most ecosystems. (A keystone species is one that has an unusually important influence on other species and the ecosystem. See the special issue on human-dominated ecosystems of *Science*, July 25, 1997).

An important example of such anthropocentrism is the common view that human impact on the environment is necessarily an unnatural disturbance. Yet all organisms affect their environment in various degrees and ways; consider the various impacts of trees, fungi, earthworms, leaf-cutter ants, bees, woodpeckers, kangaroo rats, prairie dogs, beavers, wild pigs, primates, bats, elephants, plankton, starfish, and sharks (e.g., Westbroek, 1991). However, non-human organisms that influence their environment are rarely considered a disturbance, except for exotic species introduced by humans. Usually only when an organism depletes resources and degrades the habitat beyond its natural capacities or normal period for regeneration might the term disturbance be appropriately applied. Yet it could be argued that some human societies are unnatural or even anti-nature, especially industrial ones. Other societies appear to be an integral part of nature, such as many more traditional indigenous cultures in the Amazon forests like the Yanomami. The impact of such societies on their environment is usually no less natural than that of other species. This was likely the case for the majority of societies through most of human existence.

II. BIODIVERSITY MODIFICATION

A. Extinction Hypothesis

Human antiquity markedly varies among regions, and consequently so does cumulative impact. Hominids (the human line) evolved in Africa some four to six million years before the present (B.P.). Human dispersal into other regions is approximately dated (B.P.) as follows: Europe and Asia (1 million), Australia and New Guinea (50,000–60,000), the Americas (12,000–20,000), and Pacific and other islands (30,000–to recent, depending on the island).

As humans colonized new regions, supposedly they would have a great advantage over prey that lacked previous experience with them. Relatively suddenly the hunters became a new top carnivore in the food web, rather than through the usual gradual process of predator–prey co-evolution. Furthermore, their technology provided advantages, particularly projectile weapons like the spear, that allow some safe distance from and surprise for prey when making a kill. Indeed, Paul S. Martin and others hypothesize that the dispersal of humans into new areas caused massive faunal extinctions around 50,000 B.P. or less, depending on the region, especially in Australia, the Americas, New Zealand, Oceania, and Madagascar. The main basis for the so-called "blitzkrieg hypothesis" is the apparent coincidence in timing of human arrival and massive megafaunal extinctions. However, with little to support it but circumstantial evidence, this hypothesis remains far from conclusively proven, and it is much more problematic and controversial than advocates usually admit. For example, convincing evidence is lacking in Australia for a temporal coincidence between human dispersal and megafaunal extinctions. Yet even if temporal coincidence were demonstrated, that does not automatically prove a cause–effect relationship nor exclude other potential causal factors such as climatic change. In New Zealand, however, there is no doubt that over-hunting and habitat destruction by indigenous people (Maori) caused the extinction of many animal species, including the large moa birds (see Flannery, 1995).

Whether examining faunal extinctions or any other aspect of human impact on biodiversity, human diversity must be taken into account. Even in especially vulnerable ecosystems, different cultures may have very different impacts, such as indigenous peoples and subsequent colonists in Hawaii. Hawaiians introduced 34 exotic species and supposedly caused the extinction of some 50 endemic species during their 1500 years of exclusive occupation. In contrast, European, Asian, and other colonizers introduced 4653 exotic species, caused the extinction of 211 endemic species, and endangered more than 800 others, all in a little over two centuries (e.g., Culliney, 1988). Consequently, compared to native Hawaiians, subsequent colonizers caused 137 times as many introductions and four times as many extinctions, all in a mere one-seventh of the time.

In general, traditional hunter-gatherers and others may seem to have a very limited impact on their environment, given their low population density, high mobility, limited technology, subsistence economy, minimal needs and wants, intimate environmental knowledge and monitoring, and animistic worldview, attitudes, and values that are nature oriented. Nevertheless, even small bands of nomadic foragers can have some impact, for example, on seed dispersal; one way archaeologists identify prehistoric occupation sites such as hunting camps is by distinctive combinations and concentrations of useful plant species that would not likely occur naturally.

B. Fire

Natural fires have influenced ecological systems and processes for many millions of years. In contrast, so far evidence indicates that the human use of fire extends back only about a half million years in the Old World to the later portion of the period of *Homo erectus*. Usually the main reason hunters burn an area is to create fresh plant growth to attract game. In addition, low-intensity fires set by indigenes prevent the accumulation of fuels that might otherwise lead to devastating wildfires, especially during a dry season, drought, or from lightning strikes. (This is a lesson only recently learned by national park managers in the United States, Australia, and elsewhere.) Controlled burning of different patches of the landscape at different times by indigenes or others may help increase biodiversity, whereas wildfires may decrease it.

The long-term cumulative effect of burning by foragers can be substantial, as in the case of the so-called fire-stick farming by the Aborigines of Australia over some 50,000 years (Flannery, 1995). Repeated burning of large areas of vegetation over long periods would eventually significantly modify the plant community and consequently also the associated animal community. Sometimes a pyrosere is created, a biotic community resistant or adapted to fire that is prevented from further ecological succession or development by repeated burning. With the regular and widespread use of fire by foragers, surely a new threshold in impact was reached. Some grasslands in temperate and tropical zones, such as the Scottish highlands and Venezuelan llanos, respectively, may be the result of repeated burning by humans at least in part, although in many areas this conclusion is still controversial and climatic change may be a factor as well (see S. J. Pyne, in Balée, 1998).

C. Horticulture

Fire is also an important component of farming technology. In tropical forests, during the dry season a plot of about a hectare or so of vegetation is cut and then, after drying out in the sun, it is burned. The organic ash from the burn provides fertilizer for the growth of crops that are planted just before the rainy season. This ash can be critical because tropical soils are usually poor. The crops are often pioneer or weed species that thrive in disturbed areas. Swiddening in some form is ubiquitous throughout tropical forests, but was also practiced by pioneer farmers in Europe and America.

Swiddening may enhance biodiversity. This practice creates a gap in the forest that provides sunshine for crops. They are harvested for a few years and then less intensively as productivity declines with decreasing soil fertility and increasing problems with weed competition and pests. Over years or even decades, the swidden gradually converts into fallow, with successional plants growing from seeds left in the soil and entering from surrounding forest through agencies like wind and birds. Consequently, swiddening leads to a mosaic of plant and associated animal communities at different stages of succession. Swiddening also forms ecotones (transition zones between two environments) that may increase biodiversity (edge effect) by harboring some species from both environments and others specializing on the ecotone. Such environmental heterogeneity increases the potential for higher biodiversity. Gardens and fallows also attract game and other animals, given the concentration of vegetation on the ground that provides cover and a concentration of edible plants, unlike a primary forest in the tropics.

Swiddening enhances biodiversity at the genetic level as well, since most swiddens are polycrops (numerous species and varieties) of domesticated plants. Crop diversity reduces the risk of total failure as a result of weather, disease, or pests, because some species and varieties are more resistant or resilient.

As a result of the variety of human uses and manipulations of forests, it can not simply be assumed that any are entirely pristine. In the Amazon it is estimated that nearly 12% of the forests beyond the floodplains are to some degree anthropogenic (Balée, 1994). For some two decades researchers at La Selva, the famous tropical biology research station in Costa Rica, thought that they were studying natural forest. They were unaware that it was to some extent anthropogenic until the recent discovery of charcoal and pottery fragments in the soils. In Africa, conservationists assumed that forest patches in savanna grasslands were relics of formerly more widespread forest. However, studies demonstrated that many of these forests are not natural relics of high biodiversity, but are established, used, and managed by local villagers. Thus, farming may grade into forestry with agroforestry (mixed tree crops) as an intermediate phase.

Under traditional conditions—low density of the human population, a subsistence rather than market economy, adequate fallow periods, and extensive forest for future gardens—swiddening is usually sustainable, that is, it does not irreversibly deplete natural resources and degrade ecosystems. Indeed, it does not appear that forests like the Amazon were ever endangered, even after centuries or millennia of indigenous activities, until the encroachment of Western civilization and particularly in recent decades. However, in most areas traditional conditions no longer hold.

On the other hand, there are cases where swiddening has a negative impact. As an example, changes in plant species diversity over the last 14,000 years in Panama are preserved in phytoliths (plant fossils) from lake sediments. A sharp decrease in this diversity coincides with the onset of swiddening around 7000 B.P. and may be causally related (Piperno, 1994). Thus, human impact on biodiversity from swiddening and other activities needs to be assessed on a case-by-case basis.

Still, traditional *shifting* cultivators contrast markedly with shifted cultivators of recent decades, the latter being immigrants who colonize tropical forests, often as economic refugees. Pioneer swiddeners are much more likely to contribute to deforestation because of their lack of familiarity with farming in a tropical forest and orientation to a market economy rather than mainly subsistence. This is the case with people moved by the Indonesian government in its transmigration program from islands with high human population density like Java to those with low density like Irian Jaya. A similar situation occurred when the Brazilian government built the Trans-Amazon highway as a way to relocate rural poor mainly from the northeast into the Amazon Basin.

III. BIODIVERSITY CONVERSION

A. Agriculture

The shift from human modification to the conversion of environments and biodiversity pivots on the decline of *rotational* systems of land and resource use. Traditional foragers, herders, and swiddeners usually tend to maintain sustainable economies by rotating one or more of these factors: types of resources used as well as places, times, techniques, and personnel in resource exploitation. If rotation decreases, then human impact increases markedly and eventually conversion occurs instead of modification, reflecting an emphasis on intensive rather than extensive use of land. By various estimates 30–50% of the land surface of the planet has been converted by humans.

Agriculture usually involves some kind of plow pulled by a domesticated animal, in contrast to horticulture, which usually involves only hand tools like an ax and digging stick powered solely by human muscle. Agriculture is a more intensive and permanent form of land use focused on conversion of the environment, whereas horticulture is a more extensive but temporary or rotational form of land use focused on modification of the environment. Accordingly, agriculture usually simplifies ecosystems and reduces biodiversity, whereas horticulture may sustain or even increase environmental heterogeneity and biodiversity. Agriculture also creates more environmental fragmentation than horticulture, that is, the dividing up of landscapes into isolated or semi-isolated patches of "natural" environments such as forest remnants in farm fields. However, whether horticulture or agriculture, it is necessary to assess the impact on biodiversity on a case-by-case basis.

B. Rice Paddies

Throughout Asia, large portions of the landscape, especially in the lowlands near rivers and streams, were converted centuries or millennia ago to wet rice paddies. Paddies can also be found in highlands where terracing has been developed. Conversion continues to this day, driven by increasing population and economic pressures on land and resources; it is accelerating in many areas. Conversion includes numerous types of forests, grasslands, and wetlands, and accordingly sacrifices an enormous amount of biodiversity. For instance, various types of forest covered about 70% of Thailand until World War II, whereas today only about 15% of the country is forested, largely because of agricultural expansion and logging.

Rice paddies are not simple monocrops, however. There may be hundreds or even thousands of varieties of rice in a region. It is estimated that some 50,000 local varieties of rice existed in India until recently. In addition, there may be several hundred species of wild plants and animals associated with paddies, many of them considered edible. Beyond the paddies there are home gardens, fruit orchards, pastures, various types of forest, and agroforestry (tree crops), all of which contribute considerable biodiversity to the regional landscape. For instance, in parts of Bali, home gardens may contain as many as a hundred or more species of wild and domesticated plants.

C. Raised Fields

Extensive areas of raised or ridged fields are found in many parts of the world, usually along river floodplains or in other wetlands. They elevate land above water for farming. Soil is dug and piled on top of either side of a ditch to form a long mound. The ditch becomes a canal for water drainage and irrigation. On floodplains the ridges are arranged in parallel fashion perpendicular or at an angle to the river. Organic debris and sediments from the canals are periodically dredged and placed on top of the mound as a rich organic fertilizer. Fish, waterfowl, and other aquatic organisms thrive in the canals and are harvested for protein in the diet. Raised fields may increase productivity and biodiversity beyond the level that would otherwise exist in the region.

D. Vavilov Centers

Vavilov centers are concentrations of genetic and morphological diversity remaining at the several foci of plant and animal domestication from the Neolithic some 5000–10,000 B.P., the exact period depending on the region. These centers, as concentrations of landraces and their wild ancestors, are in effect *in situ* gene banks. (Landraces are species and varieties of domesticated plants and animals that have been genetically improved by traditional farmers and herders, but remain uninfluenced by modern breeding technology.) Prehistoric and historic farmers and herders engineered significant biodiversity in their crops and livestock, in part to reduce risk. For instance, some 8 species and 3000 varieties of potatoes were cultivated in the Andes. As another example, today a single species of sheep (*Ovis aries*) includes more than 800 different breeds, many quite ancient (Heywood, 1995: 111, 775). Vavilov centers are increasingly endangered and degraded by growing population and economic pressures. Thus, attempts at *ex situ* conservation of their biodiversity may be the main hope for preserving samples. The loss of diversity in the varieties of domesticated species and their ancestors severely limits the potential for adaptive responses to future environmental changes and perturbations, whether natural or anthropogenic.

E. Population

It has been estimated that through agriculture and other activities, humans preempt some 40% of the earth's total primary biological production annually (Vitousek *et al.*, 1986). This magnitude of resource use surely diminishes the possibilities for many other species and the planet's biodiversity as a whole. Given projections for continued human population growth in this century before any global stabilization, surely this co-option of energy and nutrient sources from other species will get much worse.

The human population explosion is a relatively recent phenomenon. After some four million years of human evolution, the world population had only reached 300 million by A.D. 1000. Then world population grew to 500 million by 1500, 1 billion by 1900, and recently topped 6 billion. It is projected to be 10 billion by 2050. Correlated with this growth is increasing population density. For example, generally in the tropics human population densities (individuals/km^2) are: foragers, <1; swiddeners, dozens; and wet rice farmers, hundreds to thousands. However, though increases in the size and density of the human population certainly increase impacts on biodiversity, not all humans create the same impact, and some are grossly disproportionate. For example, upper- and middle-income consumers use resources and produce pollution at levels many times higher than lower-income consumers and the poor, whether comparing economic classes within a single society or so-called developed and less-developed countries—the core and periphery in the world economic system, respectively. The wealthy nations, as well as petroleum and pharmaceutical companies, have a special responsibility to reinvest some of their profits to help fund biodiversity studies and conservation.

IV. BIODIVERSITY COMMODIFICATION

A. Cities

Market economies and cities are largely responsible for the commodification of nature—the commercial or monetary evaluation of biota and landscapes. Because most urbanites and suburbanites do not produce their own food, they must trade or purchase it and other natural resources as imports from the farmers and peasantry working in the vast rural hinterland. Thus, commodification has been developing with urbanization for at least 5000–6000 years. Today about half of humanity lives in an urban environment, although this concentration, the megalopolis, and suburban sprawl are recent phenomena. However, it is not always easy to detect a clear boundary between urban, suburban, rural, and "natural" areas. Still, one thing is clear: urbanization will continue to increase around the world and variously impact biodiversity.

Another result of the commodification of nature, and a major strategy in biodiversity conservation that is mainly advocated by persons from the urban elite, is the economic valuation of environmental products and services, such as medicinal plants and products from tropical rain forests. Although this approach may be necessary to counteract the economic pressures for large-scale resource extraction from national governments, corporations, and market forces, this sales pitch diminishes biodiversity by ignoring its diverse nonmonetary intrinsic values.

The human body is host to biodiversity in the form of a multitude of internal and external parasites and other microscopic to small organisms that inhabit and visit it, such as bacteria, fungi, mites, lice, mosquitoes, ticks, and leeches. Likewise, human habitations and settlements, especially cities, may be viewed as hosts to a peculiar assemblage of biota, mainly commensals and parasites, such as squirrels, dogs, cats, rats, mice, pigeons, cockroaches, and flies. Though cities are mostly anthropogenic environments, they may also preserve fragments or islands of "nature" of different sizes in the form of various parks (including botanical gardens, zoos, and aquariums), home yards and gardens, indoor plants, and pets. As an example, there are about 1000 species of plants in the parks and gardens of the city of Singapore. Immigrant and ethnic neighborhoods introduce biodiversity into cities through the distinctive variety of plant species cultivated in their home gardens. In some regions where biodiversity is naturally low, human habitations and settlements may elevate biodiversity beyond prior natural levels.

The microscopic level of biodiversity cannot be ignored. Large urban populations also serve as a reservoir for viruses and a diversity of other microbes that transmit and maintain diseases. The emergence and reemergence of new infectious diseases and drug-resistant strains are other aspects of how microbial diversity and history are often linked to cities. The high density of the human population in cities not only creates sanitation and health problems for humans, but can also degrade the water quality and life in nearby aquatic ecosystems.

B. Colonialism

Colonialism, in its various forms, has affected biodiversity throughout the world, and continues to do so. During the last 500 years, Europeans introduced plants and animals from the Old World into their overseas colonies, especially areas that were climatically similar to parts of Europe, like portions of Burma, California, Chile, Kenya, and South Africa. This Europeanization of local ecosystems reduced native biota and created neo-Europes. The biological exchange also flowed in the opposite direction, however, from the New World (Americas) to the Old (Europe and beyond), with the spread of crops originally domesticated by prehistoric native Americans, such as potatoes, corn, squash, tomatoes, tobacco, cotton, and manioc. Indeed, today much of the world's food supply comes from introduced species (Crosby, 1986).

Biodiversity at the microbial level is another aspect of the so-called "Columbian exchange," the biological and cultural consequences of the Old World and New World on one another. Indeed, throughout the Americas epidemics of introduced Old World diseases caused catastrophic depopulation of indigenous communities. Because these populations had been isolated from the Old World for thousands of years, they had no previous experience with the Old World diseases to develop immunological resistance. The devastating crash of their populations relieved pressure on their land, resource base, and local biodiversity. In many areas, forests and other environments regenerated rapidly, indicating that indigenes had not irreversibly degraded them. However, the apparent demographic void and supposed wilderness encouraged European colonization of what appeared to them to be an underpopulated and underused frontier. This process continues to this day in parts of the Amazon and elsewhere (Denevan, 1992).

The Columbian exchange was certainly an unprecedented event in the history of human impact on biodiversity; however, it did not operate on any prior pristine nature. Long before European contact, the peoples of the Americas modified and in some areas converted many ecosystems (Denevan, 1992). For instance, in some areas of the ancient Mayan civilization there was extensive deforestation. As another example, anthropogenic prehistoric shell middens (mounds) in riverine and coastal zones may have higher diversity and density of useful plant species than adjacent "natural" sites.

Europeans are not the only colonials, just those most familiar. All empires are built in part on geographic expansion through some combination of long-distance trade, military conquest, and domination and oppression (economic, political, cultural, religious) for the exploitation of the land, resources, and labor of other societies. Some examples are the Aztec in Mexico; the Inca in the Andes; Arabs in many parts of Africa, Europe, and Asia; and Chinese in much of East and Southeast Asia. All empires affect local biodiversity through biotic exchanges and environmental modification and conversion. For instance, in Asia the geographical distribution of some plant species is inexplicable without considering the role of sacred species such as the bodhi or pipal tree (Ficus religiosa) in Buddhism and temple yards.

C. Modern Transportation and International Commerce

Long-distance ocean travel and trade became possible with the development of seaworthy ships some 6000 B.P. and eventually facilitated European colonialism and widespread biotic introductions. However, intentional and unintentional introductions of species between formerly independent or isolated regions of the world have greatly accelerated with the advent of modern transportation and international commerce. Now more than 600 species of animals and plants are directly threatened by illegal international trade that involves several billion dollars annually.

The introduction of exotic species can endanger native species and ecosystems. For example, the introduction of the South American otter (*Myocastor coypus*) into southern Louisiana in the 1940s had profound consequences on the landscape and ecology of the Mississippi delta. The inadvertent introduction of the brown tree snake (*Boiga irregularis*) from Papua New Guinea to the island of Guam in Micronesia caused the extirpation (local extinction) of a dozen species of native birds and still threatens other fauna. The zebra mussel (*Dreissena polymorpha*) accidentally spread from Europe into the Mississippi River basin of North America when a tanker dumped ballast waters in the Great Lakes. Among other problems, this exotic mussel now threatens numerous native mussel species. Many other examples could be given, for more than 4500 exotic species have become established in the United States during this century alone. The effect of species introductions worldwide probably results in a net loss of global biodiversity. However, in some areas, such as New Zealand and Hawaii, net biodiversity has actually increased as the result of numerous introductions of exotic species, even though they variously threaten, endanger, or extirpate many native and endemic species.

Through European colonialism, Enlightenment ideals of rationalism and individualism, industrialization, capitalism, economic development, modernization, and globalization, the world's environment, land, and natural resources, including biodiversity, have progressively become objectified and commodified. All of this has greatly facilitated the overexploitation, degradation, and destruction of biodiversity throughout the world and in its last frontiers such as tropical forests. In turn, this trend has been magnified by the insatiable greed that is inherent in the competition and growth mania of capitalist economies and the associated culture of materialism and consumerism.

This situation is compounded by the fact that the rich are getting richer at the expense of the poor who are becoming poorer. Wealthy entrepreneurs, multinational corporations, and so-called developed countries can afford to buy greater access and rights to biodiversity, just as they also exert a disproportionate impact on it. The poor are able and willing to pay less attention and money for biodiversity conservation than the rich, not necessarily because they are ignorant or unconcerned, but because they have fewer economic assets and alternatives. During the 1997 economic crisis in Southeast Asia, the harvesting and export of wildlife species and their products accelerated markedly as local people turned to their forests to generate desperately needed cash, as their regular farm crops and jobs no longer provided adequate household income. Material wealth and technology are concentrated in the developed countries in the Northern Hemisphere, whereas population growth, poverty, and biological wealth are concentrated in the so-called less-developed and developing countries of the Southern Hemisphere. Such inequities are a serious obstacle to biodiversity conservation as well as human well-being.

D. Commercial Farming Industry

The beginnings of various agribusiness industries can be traced back to European colonialism. Most of these industries cause the massive conversion of forests and other ecosystems to monocrop plantations, such as in the tropics with bananas, cocoa, coconut, coffee, cotton, eucalyptus, oil palms, pineapple, rubber trees, sugarcane, tea, and tobacco. As a result, higher thresholds of environmental impact have been reached with extensive biodiversity loss at the genetic, species, and ecosystem levels.

Since the 1960s, the so-called Green Revolution and other forces of globalization have been threatening and diminishing the genetic and species diversity of domesticated plants and animals that have been developing since the Neolithic period. Monocrops from the Green Revolution are genetically engineered for rapid growth and high productivity in response to massive subsidies of chemicals as fertilizers, herbicides, and pesticides.

However, because of the high costs of inputs, mechanization, and transportation, these agribusiness industries are grossly inefficient even if highly productive. Such monocrops are also much more vulnerable than the diverse agroecosystems of traditional farms to climatic changes and weather perturbations, such as global warming and El Niño. Nevertheless, modern agribusiness, biotechnology and genetic engineering, and bioprospecting are currently revolutionizing food production among other things, but with largely unknown yet probably far-reaching consequences for the maintenance of biodiversity.

The forests and woodlands of the world have declined significantly in area since the origin and spread of agriculture some 10,000 years ago. However, deforestation rates have surged dramatically in recent centuries, and especially in recent decades in the tropics. Although tropical forests cover only about 6% of the earth's surface, they contain about half of all its biodiversity; thus they are the recent and future hot spots for biodiversity research and conservation. These forests are also increasingly becoming the hot spots of sociopolitical conflict and violence, such as in the Democratic Republic of Congo, Rwanda, Burundi, and Sierra Leone. Many temperate forests are deteriorating from loss of old growth, acid rain damage, and other forces. However, the forest cover in New England has expanded as the farm economy declined over the last 100 years, a hopeful indication of nature's capacity for regeneration.

The combination of causes and the magnitude of deforestation and consequent biodiversity erosion have varied tremendously through time and space depending on the specific details of the particular context. The impact of a stone ax on tropical forest compared to a metal one is quite different, even allowing for cumulative impact with substantial human antiquity in an area. The introduction of more efficient metal tools and a market economy to trade local natural resources for Western manufactured goods have accelerated deforestation throughout the tropics. However, in recent decades newer technologies, including chain saws and bulldozers, together with new economic enterprises and incentives, created unprecedented rates of deforestation. In many areas of the Brazilian Amazon, for instance, cattle ranching made possible by government tax support and subsequent land speculation has been the major cause of deforestation since the 1960s. In much of Central America, forests were converted to pasture for cattle to feed beef for the fast food industry in North America and Europe, the so-called hamburger connection. Banana plantations have been another major cause of deforestation in Central America. An important factor in deforestation in many areas of Southeast Asia and elsewhere is the alienation of local communities from their land and resources by state government. Forests have been converted to monocrop plantations of eucalyptus and other fast-growing trees for export to Japan and elsewhere to supply the paper pulp industry.

Aquaculture is yet another kind of commercial and industrial farming. It has been endangering mangrove forests and other coastal ecosystems and their biodiversity throughout the tropics in recent decades. Shrimp pond farms along the southern peninsula of Thailand have grown exponentially with the investment of outside capital and technology, with the benefits going mostly to outsiders. Although aquaculture is a very ancient, productive, and efficient method for producing quality protein in many parts of Asia and elsewhere, where it has traditionally been integrated with other aspects of the economy and ecology, this new economic development is very different. The shrimp are usually much too expensive to be consumed by local populations. Instead, they are exported to distant markets in Japan, Taiwan, North America, and Europe. Yet shrimp ponds are short-lived because waste products accumulate in the sediments until they become toxic to the shrimp or for any other subsequent uses. Consequently, repeatedly aging ponds are abandoned and new ones are constructed. At the same time, this cancerous growth of shrimp farms is degrading and destroying the local economies and biodiversity of many coastal regions.

V. BIODIVERSITY TOXIFICATION

A. Industry

At least since 1962, with the publication of Rachel Carson's classic book *Silent Spring*, there has been growing concern about the environmental impact of chemicals and pollution. Because in both ecosystems and the biosphere, ultimately everything is connected to everything else in some way, chemicals from agriculture, factories, cars, and other sources eventually circulate worldwide. Residues of DDT and other pollutants are even found deposited in layers of polar ice. The accumulation of nitrates and other nutrients from agricultural and household chemicals can generate the explosive growth of algae in lakes and rivers to the detriment of biodiversity, a process called eutrophication. Massive kills of birds, fish, and other species have been reported from this and other contaminants such as oil spills.

One recent possible symptom of widespread pollution, whether direct through water and soil contamination or indirect through increased ultraviolet radiation from the ozone hole, may be the gross deformities observed in frogs, a most alarming phenomenon and perhaps an early warning of impending ecological catastrophe. Another result of atmospheric pollutants such as automobile exhaust is the greenhouse effect, which in the twenty-first century may trigger global warming of a few degrees and the consequent rise of sea level by up to two meters or more. The consequences of sea level rise on the biodiversity of coastal and marine ecosystems like wetlands and coral reefs are uncertain, but are likely to be negative and potentially catastrophic. In terms of high biodiversity, coral reefs are the marine analog of tropical rain forests.

B. Oil

Industrial society, whose lifeblood is oil and other fossil fuels that are actually the remains of ancient biodiversity, has produced unprecedented types and levels of chemical pollution that undoubtedly endanger and erode living biodiversity. To illustrate, in recent decades many remaining frontier zones like the Amazon of Ecuador and Peru have been the target for oil exploration and extraction. Because of the isolation of these frontiers, the usual environmental safeguards and measures for clean-up are ignored.

During the exploration phase, hundreds of miles of roads and large grids of extensive trails for seismic testing are cut into the forest. The seismic explosions not only scatter wildlife, but the resulting shock waves can kill hundreds of fish in rivers, lakes, and wetlands. During the production phase, a single oil well platform consumes about six acres of forest and about 2000 trees. (In the rain forest a single giant tree may be inhabited by thousands of insect and other species, most unknown to science.) Adjacent production wastes and treatment chemicals amount to millions of gallons every day for decades. At well sites there is no proper disposal of toxic waste, only open-air pits that eventually overflow into the soils, groundwaters, and surface waters, thereby contaminating and killing fish and wildlife. One gallon of oil can kill the fish living in a million gallons of water and adversely affect aquatic life at concentrations as low as one part per hundred billion.

Over decades, many hundreds of oil wells and extensive pipelines allow for numerous leaks and spills of black crude when breaks occur through metal aging or earthquakes. The magnitude of pollution that has been occurring in frontiers like the Amazon makes the oil spills associated with the Exxon *Valdez* and Gulf War look like minor irritants in comparison! The costs to biodiversity are incalculable. Furthermore, beyond the corporate irresponsibility and massive environmental destruction, in the Amazon oil is even being pursued in wildlife reserves and other areas supposedly set aside for conservation by national governments (Kimerling, 1991).

C. Militarization

A grossly neglected type of human impact on biodiversity and the environment are military activities and warfare, even though armament production is the top industry in the world with more than $800 billion in sales annually. One aspect of warfare that is especially detrimental to biodiversity is the use of scorched earth tactics, which unfortunately are nothing new. One of the largest applications of this tactic was the U.S. military's dumping of some 13 million gallons of Agent Orange to defoliate forests in the Vietnam War during 1962–1971. In recent decades, the U.S. war on drug production in Amazon forests has also involved the use of defoliants. During the Gulf War, the government of Iraq set fire to oil wells in the desert of Kuwait and created oil spills along the coasts. The tolls of such tactics on local and regional biodiversity have yet to be fully revealed.

The rise of nationalism and ethnic conflicts with the end of the Cold War has spread militarization into frontier, border, and other zones. Conflicts and wars are likely to proliferate in the future, according to Malthusian pessimists and others who consider growing resource scarcity and competition as major contributing causes. If so, then national and international security would be well served by redirecting a significant portion of funds that now go to military and defense into biodiversity and environmental studies and conservation programs.

Whereas war may threaten and erode biodiversity, peace may promote it. For instance, the Demilitarized Zone (DMZ) between North Korea and South Korea is a corridor 4 kilometers wide and 250 kilometers long that extends across the peninsula. For nearly five decades this corridor has been rigidly enforced as a no-man's-land. As a consequence, farmlands thousands of years old and degraded forests have both reverted to a natural condition, thus protecting threatened and endangered species of plants and animals, as well as a cross section of the ecosystems of the Korean peninsula

(Kim, 1997). The international peace parks between Costa Rica and its neighbors, Panama and Nicaragua, also promote biodiversity conservation by acting as refuges from human activities.

VI. BIODIVERSITY FUTURES

A. The Certain Future

As those from the "green" sociopolitical movement and many others recognize, industrial society, capitalism, and economic development are based on the *false assumption that infinite growth is possible on a finite base.* Here growth refers to both population and economy. Base refers essentially to carrying capacity. The latter includes not only the ability of the land and natural resources to support a certain level of population without resource depletion and environmental degradation, but also the capacity of ecosystems and the planet as a whole to absorb pollution and other anthropogenic stresses. Increasingly human impacts exceed the resilience of nature to regenerate and recover within a normal time period from any perturbations and stresses, natural or anthropogenic.

Until the demonstrably ecocidal ideas and practices of modern industrial society and related factors are corrected and ecosanity with some modicum of ecological balance is restored, the net impact of humans on biodiversity will be negative. Accordingly, all life, including that of humanity, will remain endangered. There isn't much room for optimism, given the great momentum of population and economic growth combined with political pressures for so-called economic development and the elevation of the standard of living throughout the world, all at the expense of the environment and its natural capital of resources, biodiversity, and ecosystem services. Indeed, the ozone hole, greenhouse warming, acid rain, collapse of oceanic and other fisheries, soil erosion, desertification, and other global environmental problems may be symptoms of the failure of the experiment of industrial society after just two centuries. No human society is infallible and eternal—the archaeological and historical records provide many examples of those that became maladaptive, collapsed, and disintegrated like Rapa Nui (Easter Island), Harappa in the Indus Valley, or ancient Greece and Rome (Ponting, 1991; Redman, 1999).

B. Limits of Government

National and international governmental and non-governmental agencies have been making some significant progress on resolving environmental problems and managing the human impact on biodiversity. Nevertheless, there are many serious limitations on the efficacy of government-protected areas for biodiversity conservation. First, they comprise only about 5% of the earth's surface, which is not a very large or representative sample of the tremendous biodiversity of the planet. Second, they are often little more than "paper parks" because of inadequate funding and administration. Third, they will come under increasing attack in many ways with accelerating population and economic pressure in the future. Fourth, even for the some 10,000 sites that are supposedly protected by governments, only 5% of these have been thoroughly inventoried for biodiversity, and it would take the equivalent of the current number of experts several centuries to inventory the remainder.

Clearly there is an enormous challenge for inventorying, managing, and conserving biodiversity. Indigenous and other local communities can make a significant contribution to biodiversity studies and conservation, and thereby also have a positive impact on biodiversity. This is just beginning to be appreciated; an example is INBio (Instituto Nacional de Biodiversidad), the biodiversity inventory and conservation program in Costa Rica that employs local people as parataxonomists. Also, though today there are some 1600 botanic gardens in the world that help conserve plant diversity *ex situ*, there are nearly 7000 distinct cultures in the world and each may somehow contribute to *ex situ* and *in situ* conservation. Comanagement is usually the ideal, that is, the cooperative sharing between community and government in the design, authority, responsibility, and benefits of natural resource management and biodiversity conservation projects. Among the successes in comanagement are Manu National Park in the Peruvian Amazon and Kakadu National Park in northern Australia. However, when government administrators of protected areas ignore the needs of local people, then conservation efforts usually falter or fail.

Numerous people have grown increasingly skeptical that government, science, technology, and education are sufficient for resolving the spiraling environmental crises facing the world. They view such attempts as treating superficial symptoms of the crises, not the broader underlying causes. Some of these people are turning to their own religion as a source of worldviews, inspiration, motivation, attitudes, and values for developing a more sustainable and meaningful relationship with nature. As the pioneers in spiritual ecology, indigenes can also provide profound insights for such endeavors.

C. Indigenous Potential

Indigenous cultures contrast sharply with industrial and other "modern" cultures. Many indigenous societies, especially those that are more traditional, may provide heuristic models for biodiversity conservation through their intimate environmental knowledge (ethnoecology), sustainable economy, natural resource management and conservation practices, spiritual ecology, and protection of sacred places. (To a large extent this is because most indigenes lie to the left on the continua described in Table I.) In these and other respects, it could be argued that many indigenous societies are actually more developed than industrial ones!

Regarding traditional environmental knowledge (TEK), for example, the Ka'apor people in the Brazilian Amazon recognize at least 768 species of plants from seed to reproductive adult stages. This reflects a tremendous amount and depth of knowledge about the biodiversity in their habitat, and that example is just in their domain of useful plants (Balée, 1994). Indeed, such knowledge of indigenous and other peoples may provide biologists and conservationists with one desperately needed short-cut for the inventory and conservation of local biodiversity.

Indigenes often consider their environment, including the biotic and abiotic components, to be sacred. Such a worldview can encompass a deep respect and reverence for nature that tempers their cultural ecology and resource use, and that may lead to environmental conservation, inadvertently if not intentionally. The role of spiritual ecology and sacred places in biodiversity conservation is just beginning to be recognized and explored by those who have an open mind to such phenomena.

Most indigenous and numerous other religions are nature centered, considering certain areas of "nature" in their habitat to be the foci of spiritual power. These sacred places are treated with extraordinary care and respect, and frequently taboos restricting resource use are associated with them. However, throughout the world aggressive Christian missionization and other dominating monotheistic religions have often destroyed sacred places and indigenous religions because they were perceived as pagan; these religion-inspired actions threatened and eroded biodiversity as well. The objectification and commodification of nature also degrade and destroy sacred places, as Australian Aborigines have found with Western mining activities. Nevertheless, numerous and diverse sacred places in nature remain and may have contributed to biodiversity conservation in the past and/or could do so in the future.

D. Conservationists and/or Exterminators

The preceding discussion does not support the romantic idea that *all* indigenes are *always* in *perfect* harmony with their environment, the myth of the so-called "ecologically noble savage," which amounts to little more than a "straw man" argument. Instead, it recognizes the fact that many indigenous societies were to some degree environmentally friendly and that some still are, even though others are not or were not. These variants of the human impact on biodiversity have been documented in numerous cases by cultural and historical ecologists. Revisionist advocates who attack indigenous societies as environmentally destructive have yet to realize, let alone adequately resolve, the basic contradiction in their argument—how indigenes can be so knowledgeable about their habitat and interact with and monitor it on a daily basis, yet be so ignorant of, or amenable to, such destructive practices.

A related fallacy is that all humans are environmentally destructive, the so-called "*Homo devastans*" or "humans as the exterminator species" view (Balée, 1998). Such simplistic either–or, all-or-nothing, always-or-never thinking is misleading at best, but it is a surprisingly common deficiency of advocates of this view. Blaming all of humankind for a negative impact on biodiversity is simply scientifically inaccurate, sloppy scholarship, and professionally irresponsible. It needs to be emphasized again that humanity is diverse and so is its impact on biodiversity—some societies decrease it, others sustain it, some enhance it, and others affect it in some combination of these directions.

E. Diversity Principle

In general, several authorities have observed independently that the greatest concentrations of biological diversity tend to coincide with those of cultural diversity, especially in tropical forest areas, and most of all in the so-called megadiversity countries of Brazil, Colombia, Mexico, Congo, Madagascar, Indonesia, and Papua New Guinea. The present author calls this general tendency toward a geographical coincidence of high cultural diversity and high biodiversity the *diversity principle*. It has barely begun to be recognized and systematically described, let alone explained. However, wherever indigenous societies thrive, biodiversity is likely to do so as well.

One thing is certain: in such megadiversity regions, threats to either cultural or biological diversity also threaten the other. Ironically, many indigenous societies that have proven sustainable and adaptive for centuries or even millennia, as well as their environments, are being degraded and destroyed by industrial and other societies that have yet to stand the test of time and increasingly show clear symptoms of maladaptation. The concerns for biodiversity conservation and human rights are interdependent. Furthermore, the degradation and destruction of cultural diversity, like that of biodiversity, seriously endanger the future adaptability of *Homo sapiens* as well as biological evolution in general.

VII. CONCLUSIONS

Globally the net impact of the human species has most likely decreased biodiversity. However, *not all humans are equal in their impact on biodiversity because of the tremendous diversity in humankind throughout its temporal and spatial distribution including cultural diversity.* At the population level, clearly some societies may sustain or even enhance biodiversity. In particular, many indigenous societies, especially those who retain some core traditions despite superficial changes, have special potential in their environmental knowledge, worldviews, and other attributes to contribute to developing systems for the sustainable use, management, and conservation of biodiversity. The cumulative and collective impact of humans on biodiversity across the world is sufficient to make it imperative that *anyone concerned with the biodiversity of any area must consider the possibilities of human influence.* Accordingly, research on cultural ecology and historical ecology is indispensable.

Current anthropogenic extinction rates are estimated at 1000 to 10,000 times higher than normal background rates. Furthermore, these recent anthropogenic extinctions also involve plants, whereas prehistoric extinctions mainly affected animals. This is an alarming fact, among other reasons, considering how fundamental plants are to other life as primary producers in capturing solar energy through photosynthesis. Another distinction of the present extinction spasm is that increasingly humans are becoming aware of what they are doing and could change their behavior to reduce their negative impacts. After all, destroying biodiversity and ecosystems is ultimately ecocidal for humanity since they are our life-support systems.

Biodiversity is unlikely to be adequately conserved only by preservationism—by isolating nature from human "disturbance" in a few areas of supposed wilderness. Much more effort needs to be directed to adequately recognizing and better managing human impacts from the local to the global levels and over the long-term. One problem is that decisions and actions that seem reasonable in the short-term may have negative consequences in the long-term. Thus, adequate environmental and resource management requires, among other things, much better informed politicians and policy-makers at all levels (local to international) who have the political will and morality to consider no less than the integral relationship between humanity and biodiversity for many generations to come. Of course, so far such leadership is grossly inadequate, but not unprecedented. For instance, the Iroquois in North America acted with the seventh generation into the future in mind. Perhaps the Convention on Biological Diversity from the 1992 Rio Summit is a hopeful change.

Biodiversity conservation also depends on a much more informed, concerned, and involved public that understands the nature and consequences of human impact on biodiversity. Considering the gravity and urgency of this subject, environmental and biodiversity education must be advanced at all levels and include mass media. In the process, environmental ethics must be first and foremost. One of the best places to begin is by exploring Aldo Leopold's (1949: 262) land ethic: "A thing is right when it tends to preserve the integrity, stability, and beauty of the biotic community. It is wrong when it tends otherwise." (However, today the term "resilience" is more appropriate than "stability.") Every individual decision potentially has some impact on biodiversity, however small or indirect; and those of humanity collectively can be synergetic and life threatening. (For teaching, especially useful videos are: "World Population," "Biodiversity," "Web of Life," the series "Living Planet," "Can the Tropical Rainforests Be Saved?," "An Ecology of Mind" from the Millennium series, and "Spirit and Nature.")

Finally, because the human impact on biodiversity both locally and regionally is mixed with positive as well as negative influences, there is reason for hope as well as despair for the future. However, with the current magnitude of biodiversity loss and other environmental problems, there is no doubt that the very viability, resilience, and habitability of too many of the world's ecosystems, and thereby the biosphere as a whole, are at risk, as are the very futures of organic and human evolution on this planet. Globally the net impact of *Homo sapiens* in reducing biodiversity is a dangerous reversal of the megatrends that held throughout the last 3.75 billion years of organic evolution on earth—namely, increasing diversity, complexity, and adaptability. Furthermore, recovery from this extinction spasm may require many millions of years. Because of the gravity and urgency of the negative impact of humans on the earth's life, this impact is perhaps the single greatest challenge facing humankind and biodiversity studies for the twenty-first century.

See Also the Following Articles

COMMONS, THEORY AND CONCEPT OF • ECOLOGICAL FOOTPRINT, CONCEPT OF • ECOSYSTEM SERVICES, CONCEPT OF • EXTINCTIONS, MODERN INSTANCES OF • HISTORICAL AWARENESS OF BIODIVERSITY • INTRODUCED SPECIES, EFFECT AND DISTRIBUTION • KEYSTONE SPECIES • POLLUTION, OVERVIEW

Bibliography

Agrawal, A. (1995). Dismantling the divide between indigenous and scientific knowledge. *Development and Change* 26, 413–439.

Balée, W. (1994). *Footprints of the Forest: Ka'apor Ethnobotany—The Historical Ecology of Plant Utilization by an Amazonian People.* Columbia University Press, New York.

Balée, W. (ed.). (1998). *Advances in Historical Ecology.* Columbia University Press, New York.

Berkes, F. (1999). *Sacred Ecology: Traditional Ecological Knowledge and Resource Management.* Taylor & Francis, Philadelphia.

Bodley, J. H. (2000). *Cultural Anthropology: Tribes, States, and the Global System*, 3rd ed. Mayfield Publishing, Mountain View, California.

Crosby, A. W. (1986). *Ecological Imperialism: The Biological Expansion of Europe, 900–1900.* Cambridge University Press, Cambridge, United Kingdom.

Culliney, J. (1988). *Islands in a Far Sea: Nature and Man in Hawaii.* Sierra Club, San Francisco.

Denevan, W. M. (1992). The pristine myth: The landscape of the Americas in 1492. *Ann. Assoc. Amer. Geographers* 82, 369–385.

Fairhead, J., and Leach, M. (1996). *Misreading the African Landscape: Society and Ecology in a Forest–Savanna Mosaic.* Cambridge University Press, New York.

Flannery, T. F. (1995). *Future Eaters: An Ecological History of the Australasian Lands and People.* George Braziller, New York.

Harmon, D. (1996). Losing species, losing languages: Connections between biological and linguistic diversity. *Southwestern J. Linguistics* 15(1–2), 89–108.

Heywood, V. H. (ed.). (1995). *Global Biodiversity Assessment.* Cambridge University Press, New York.

International Union for Conservation of Nature and Natural Resources. (1997). *Indigenous Peoples and Sustainability: Cases and Actions.* International Books, Utrecht, Netherlands.

Kim, K. C. (1997). Preserving biodiversity in Korea's demilitarized zone. *Science* 278, 242–243.

Kimerling, J. (1991). *Amazon Crude.* Green Ink, Inc., Washington, D. C.

Kinsley, D. (1995). *Ecology and Religion: Ecological Spirituality in Cross-Cultural Perspective.* Prentice–Hall, Englewood Cliffs, New Jersey.

Leopold, A. (1949). *A Sand County Almanac.* Oxford University Press, New York.

McNeely, J. A. (ed.). (1995). *Expanding Partnerships in Conservation.* Island Press, Washington, D. C.

Oelschlaeger, M. (1991). *The Idea of Wilderness: From Prehistory to the Age of Ecology.* Yale University Press, New Haven, Connecticut.

Oldfield, M. L., and Alcorn, J. B. (eds.). (1991). *Biodiversity: Culture, Conservation, and Ecodevelopment.* Westview Press, Boulder, Colorado.

Piperno, D. R. (1994). Phytolith and charcoal evidence for prehistoric slash-and-burn agriculture in the Darien rainforest of Panama. *The Holocene* 4(3), 34–50.

Ponting, C. (1991). *A Green History of the World: The Environment and the Collapse of Great Civilizations.* Penguin, New York.

Ramakrishnan, P. S., Saxena, K. G., and Chandrashekara, U. M. (eds.). (1998). *Conserving the Sacred for Biodiversity Management.* Science Publishers, Enfield, New Hampshire.

Redford, K. H., and Mansour, J. A. (eds.). (1996). *Traditional Peoples and Biodiversity Conservation in Large Tropical Landscapes.* American Verde Publications, Arlington, Virginia.

Redman, C. L. (1999). *Human Impact on Ancient Environments.* University of Arizona Press, Tucson.

Soulé, M. E., and Lease, G. (eds). (1994). *Reinventing Nature? Responses to Postmodern Deconstructionism.* Island Press, Washington, D. C.

Sponsel, L. E., Headland, T. N., and Bailey, R. C. (eds.). (1996). *Tropical Deforestation: The Human Dimension.* Columbia University Press, New York.

Stevens, S. (ed.). (1997). *Conservation through Cultural Survival: Indigenous Peoples and Protected Areas.* Island Press, Washington, D. C.

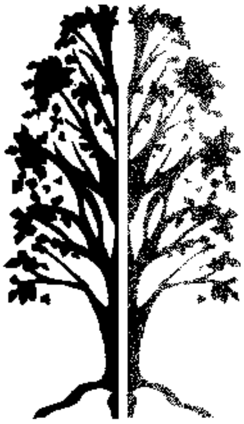
Turner, B. L. (ed.). (1990). *The Earth as Transformed by Human Action.* University of Chicago Press, Chicago.

Ventocilla, J., Herrera, H., and Nunez, V. (1995). *Plants and Animals in the Life of the Kuna.* University of Texas Press, Austin.

Vitousek, P. M., Ehrlich, P. R., Ehrlich, A. H., and Matson, P. A. (1986). Human appropriation of the products of photosynthesis. *BioScience* 36(6), 368–373.

Westbroek, P. (1991). *Life as a Geological Force: Dynamics of the Earth.* W. W. Norton & Co., New York.

Wilcox, B. A., and Duin, K. N. (1995). Indigenous cultural diversity and biological diversity: Overlapping values of Latin American ecoregions. *Cultural Survival Quart.* 18(4), 49–53.



HUNTER-GATHERER SOCIETIES, ECOLOGICAL IMPACT OF

Kathleen A. Galvin
Colorado State University

GLOSSARY

bands	The basic economic, social, and political unit of hunter-gatherer societies.
exogamy	The practice of a person seeking a mate outside of his or her group.
patrilocal residence	The practice of married couple's living in the husband's community.

A HUNTER-GATHERER OR FORAGING SOCIETY is a group of people whose subsistence is based on the hunting (or fishing) of animals and gathering of plants. Whether or not foragers have an impact on their environment depends on several factors, some of which emanate from foragers themselves and others which are external to their society.

I. INTRODUCTION

Many people have impressions of hunter-gatherers as people who live in harmony with nature, who are organized into simple societies and are associated with our "pristine" paleolithic hunter-gatherer past. Many of these stereotypic impressions are false (cf. Moran, 1991; Wilmsen, 1989). Today all foragers live in nation-states, have some dependence on either crop cultivation or farmers, and are not isolated. Hunter-gatherer societies have social systems that are extremely complex and whose interactions with the biodiversity surrounding them are as complicated and variable as was probably the case 10,000 years ago when all humans were foragers. It is no accident that today, areas with the greatest remaining biodiversity are also the areas inhabited by hunter-gatherers. Many hunter-gatherers retreating from land appropriation, settler immigration, and European diseases have occupied the most remote parts of their region. Today, these homelands are often part of or adjacent to conservation areas, parks or other protected areas.

This chapter describes traditional hunter-gatherer societies and the adaptations these societies have made to the environment. However, since hunter-gatherer societies and their environments have undergone continuing changes, issues of biodiversity conservation and hunter-gatherer welfare are discussed with the context of their changing world.

II. HUNTER-GATHERER SOCIETIES AND NATURAL RESOURCE EXPLOITATION

Because hunter-gatherers live in diverse environments they manifest an incredible diversity of cultures and natural resource management adaptations. Nevertheless, there are several general characteristics of hunter-gatherer societies; these traits have a direct impact on the use of natural resources. Traditional hunter-gatherer societies are comprised of bands, social groups made up of close biological kin and friends. The composition and sizes of bands change seasonally, depending on the abundance and location of food resources. Bands are lead by individual hunters who are respected for particular talents such as singing or dancing well, good storytelling, or hunting prowess. Other features of band organization are small group size, flexible but primarily patrilocal residence, and strong pair bonds between individual men and women. Marriage is exogamous, that is, females are recruited from other groups. These features of hunter-gatherer society are a reflection of ecological, economic, and social necessity. For example, Efe Pygmy hunter-gatherer men of the Ituri Forest in the former Zaire have very strong relationships with close kin, which facilitates defense of their territories against other cooperative kin groups. Moreover, related men can assure women access to valuable resources in neighboring Lese agricultural villages. And women are attracted to men who can guarantee long-standing reciprocal economic relationships with Lese villages. Competition for women is high so close relations with kin may also help to obtain marriageable women and provide protection, as some women leave Efe society to live in Lese agricultural villages.

Hunter-gatherers are sedentary or nomadic depending on the distribution and dynamics of their resource base. Typically, men hunt and fish while women gather and collect foods. Sometimes women's work contributes more to the diet and sometimes male hunting and fishing products are most important. Gathering of wild foods tends to contribute more to the diet among people inhabiting tropical and semitropical areas (e.g., San Bushmen of the Kalahari) than in northern temperate climates (e.g., the Inuit of Canada) where hunting contributes the bulk of the diet.

Foragers learn about their environment and resource use through acculturation. Parents teach their children different kinds of ecological knowledge and resource exploitation strategies.

Ecological knowledge is a source of landscape manipulation. For example, the Kayapo Indians of Brazil create forest islands of planted semidomesticated crops of medicinal species, wild yams, and bush bean, as well as domesticated plants such as taro, papaya, and banana. A fully grown island has sites that vary in shade and moisture thereby creating the opportunity for cultivation of different crops. They become, through time, forest patches of varying successional stages within the savanna. Cree Indians of North America rotate their hunting and fishing lands yearly to reduce wildlife disturbance and increase harvests. Biodiversity conservation is, in this case, an indirect effect of resource management. There is evidence that until recently Indians of Canada used fire to maintain trails and to open up meadows. This provided improved habitat for ungulates and increased hunting success. Australian aborigines used fire to clear trails (of poisonous snakes) and keep game habitat open.

Appropriate use of natural resources are maintained through moral and belief systems of forager societies, which includes a strong respect for nature. Through religious belief and social conventions, people respect and exert some control their natural resources. These beliefs, however, do not always prevent hunter-gatherers from overusing their resource base. Not all hunter-gatherers live always harmoniously with the environment. Indeed, evidence of escalating overuse is accumulating (e.g., Redford and Mansour, 1996).

III. CONSERVATION AMONG HUNTER-GATHERERS

It has been suggested that the hunter-gatherer adaptation occurred in environments where resources were freely available to all and were abundant. Thus, the environment was one where subsistence strategies emphasized short-term returns over long-term conservation. But during the Neolithic rise of agriculture, natural ecosystems were compressed and the value of resources increased as relative abundance declined. Some scholars have suggested that self-regulatory mechanisms evolved under resource limitation in some hunter-gatherer societies (Berkes and Folke, 1998).

There has been much written about how hunter-gatherers are actively engaged in conserving resources, especially animal resources. However, the limited actual data gathered on the subject suggests that subsistence hunters do not conserve prey resources. Most work shows that hunters are concerned about short-term gains and not about resource conservation. Small, mobile groups may use resources in a sustainable manner, for example, by maintaining small groups and ranging over a large territory, but this does not necessarily imply they are consciously conserving resources. Evidence suggests that some resources may be used intensively or even depleted in local areas while other resources are sparingly used. For example, Alvard (1998) has shown that the Piro hunters of Peru depleted the large primates in the area around their village yet have not done so to peccaries. Likewise he shows that the Indonesian Wana have nearly depleted their area of macaques (large primates) but hunt pigs in a sustainable manner. These and other studies (e.g., among the Inuit of Canada, the Ache of Paraguay, the Cree of Canada) show that both overexploitation and conservation may be practiced by hunting groups. But the point remains that hunters sometimes reduce prey species to the point of local extinction.

One plausible explanation for resource depletion is that the resources exploited by subsistence hunters are considered to be open-access resources. Open access implies that there are no controls over resource use, which is said to result in the "tragedy of the commons" (Hardin, 1968). This concept proposed that deterioration of open-access grazing land is inevitable when individuals see no benefits from resource conservation. Another reason for resource depletion is lack of concern for very abundant resources. Some level of scarcity adds value to a resource relative to when resources are quite abundant. Resource users are motivated to conserve only when they see benefits to nonuse of resources. Thus, it is only when long-term benefits outweigh the short-term benefits that conservation is expected. When tied to a specific resource base and well-defined territories, hunter-gatherers have long-term strategies for natural resource conservation (Alvard 1998). For example, traditional Maine lobstermen have strong norms of territory ownership, which are enforced through threats of violence and damage to property.

Although foragers may or may not overuse resources, their perception of the land and its value is based on use rights. Local biological diversity is an important element of local survival strategies. This view contrasts with the western view of biodiversity conservation, which is based in Western epistemology. In the western view, nature exists apart from humankind and has value independent of human use. Biodiversity conservation implies no resource use or restraints in resource use.

IV. MODERNIZATION PROCESSES AND HUNTER-GATHERERS

Major changes in hunter-gatherer society are occurring even in the most remote regions of the world. These changes are associated with agricultural development, infrastructure advancement, resettlement schemes, tree harvesting, mining and oil exploration, and other types of development. The building of roads makes it easier for outsiders to gain access to remote areas and the resources therein. In addition, hunter-gatherer populations are growing, altering their relationship to the land. The result is that indigenous systems of resource use are changing due to both internal and external pressures. The traditional systems of resource use are less appropriate or are sometimes ineffectual under current conditions. For example, traditional sanctions to protect or at least not exhaust resources are becoming ineffectual as cash income has become increasingly important to individuals interested in commodities from the modern world. Hunter-gatherers now have, under these conditions, a growing demand for cash and market goods. Under these conditions, it is less likely that people will give priority to conservation.

V. LAND TENURE, INSTITUTIONS, AND BIODIVERSITY

One political factor that is almost universally common among hunter-gatherers today is that they do not control the land they live on. Until recently, their remoteness meant that they and the resources on which they depended were somewhat protected from outside influences. Thus, resources were locally controlled by informal norms through individual behavior. Now, however, national governments, among others, have put native lands to "productive" use. This means that if the market for some product is strong it will be exploited or cultivated regardless of environmental impact. For example, the strong local demand for aguaja (a local plant) in the Peruvian Amazon has led to destructive harvesting. In theory, most hunter-gatherer communities have use rights to their territories but old laws and treaties are continually violated. Legalizing communal resource-use rights is a way of giving hunter-gatherers a long-term stake in conserving the resources on which they depend. Securing rights to resources can occur through various management and development institutions. This means that hunter-gatherers, who formally did not have institutions for collective action in the formal sense, find the need to deal with western institutions to acquire control over their lands.

The future of biodiversity, conservation, and hunter-gatherer sustainability depends on understanding that there are fundamental differences in the concept of conservation for westerners and for indigenous hunter-gatherers. Understanding that there are different worldviews toward nature is fundamental to forming a relationship between outside conservation groups and hunter-gatherer peoples. The reality is that even if hunter-gatherers are using resources, selling wild animals and cutting down trees, they perhaps remain the most effective conservationists for their region. Therefore, acceptance that there are different ways of viewing the world is a first prerequisite to working with indigenous hunter-gatherer populations. Second, it is necessary to recognize that there are no longer any "pristine" hunter-gatherers and they have needs just like the rest of us. Third, securing land tenure for hunter-gatherers and biodiversity conservation is required for a basis of a "sustainable" interaction.

VI. ECONOMIC DEVELOPMENT AND BIODIVERSITY CONSERVATION

Community-based conservation is a concept aimed at involving local people in the conservation of wildlife or protection of biodiversity. The concept developed from the realization that much of the planet's wildlife and biodiversity exist outside protected areas and in regions occupied by rural people in developing countries. Models of community-based conservation adhere to the notion that if local communities can derive some value, nominally income, through conserving biodiversity, they will do so. This promising concept has been widely promoted as "the answer" to conservation in developing countries. Thus, several models of community-based conservation have developed. The biosphere reserve is one kind of conservation area that theoretically allows for local population involvement in management of the protected areas. Integrated Conservation-Development projects are another type of community-based development. However, results from community-based conservation projects in Africa and elsewhere suggest that there are more failures than successes. Many community-based conservation efforts involve local communities in name only. Locals are neither involved in project identification and planning nor are they beneficiaries, thus these projects are not really community-based conservation projects. Another pattern of failure includes involvement of the local people only in a cursory way. Other scenarios for failure also have in common insufficient involvement of the local people at all levels in the project. In order for community-based conservation to work, people need to be considered a component of the ecosystem being conserved and brought into the project process from the beginning.

A. It Is Useful to View Humans as Part of Ecosystems

One of the fundamental problems with community-based conservation is that hunter-gatherers as well as other indigenous populations are often viewed as an external disturbance to the natural system rather than as integral components of the ecosystem. But hunter-gatherer societies see their relationship with the environment as one; they are part of that environment. Though not a study of foragers, but rather herders who do some hunting and gathering, the South Turkana Ecosystem Project (Ellis and Swift, 1988; Little and Leslie, 1999) is one of the only truly interdisciplinary and long-term projects to study the social behavior, knowledge systems, demography, human biology, and ecology of a group of people. An important goal of this study was to understand how the environment affected human management and how people affected the environment. In this case people and livestock (camels, cattle, sheep, goats, and donkeys) lived in a harsh, dry, and highly seasonal environment. This assemblage of people, livestock, plants, and other organisms within a semiarid ecosystem produced a remarkably interactive system.

Vegetation structure in this tropical savanna and dry woodlands was shown to be hierarchically constrained by physical factors: by climate at regional scales, by topography and geomorphology at landscape scales, and by water redistribution and disturbance at local and patch scales; livestock and humans played a small role. The pastoralists did influence vegetation composition and cover by burning, woodcutting, and through seed distribution by livestock. These influences were small. Livestock ecology and production followed those of the seasonal dynamics of plants. The different patterns of forage utilization by different herbivores, plus differential habitat use, lead to almost complete niche separation among this suite of domestic herbivores; among all five species, they managed to utilize a wide variety of the available plant types in the ecosytsem. Thus, physical heterogeneity on the Turkana landscapes ultimately resulted in spatial and temporal variation in plant production, plant life form diversity, and refuge areas for pastoralists These, in turn, contributed to social and ecological persistence by reducing variability of ecosystem energy flow and long-term variations in species diversity. Thus, biodiversity was important to ecosystem (which included people) maintenance. This systems approach to understanding human-environment interactions is a useful way to discern the ecological impact of hunter-gatherers and, more important, to derive appropriate management of lands where hunter-gatherers live.

The description presented here shows that indigenous concepts of conservation, ecological knowledge, and moral and religious beliefs are fundamental to understanding how hunter-gatherers use resources. Not all hunter-gatherers conserve their resources, thus whether or not and to what extent hunter-gatherers effect their environment is an empirical question that needs to be investigated, not a notion to be assumed one way or another. It is, however, the case that when hunter-gatherers have short-term strategies for resource use they may overuse some resources; when long-term goals are in place, they do not. Informal institutions control use of some resources in hunter-gatherer societies, but collective action or formal institutions are generally not well developed. With major changes in and around the lands inhabited by hunter-gatherers, it is becoming increasingly necessary for hunter-gatherers to develop institutions to gain control over their resource base. Alliances between hunter-gatherers and others interested in conservation may facilitate resource-management strategies that reduce the impact of negative changes. Hunter-gatherer natural resource-management strategies that include their social system are important attributes of these ecological systems and need to be fundamental components of any plan to conserve biodiversity.

See Also the Following Articles

BIODIVERSITY-RICH COUNTRIES • ETHNOBIOLOGY AND ETHNOECOLOGY • INDIGENOUS PEOPLES, BIODIVERSITY AND • LAND-USE PATTERNS, HISTORIC • RELIGIOUS TRADITIONS AND BIODIVERSITY • TRADITIONAL CONSERVATION PRACTICES

Bibliography

Alvard, M. S. (1998). Evolutionary ecology and resource conservation. *Evolutionary Anthropology* 7(2), 62–74.

Bailey, R. C., and Aunger, R., Jr. (1989). Significance of the social relationships of Efe Pygmy men in the Ituri Forest, Zaire. *American Journal of Physical Anthropology*. 78, 495–507.

Berkes, F., and Folke, C. (1998). Linking Social and Ecological Systems. *Management Practices and Social Mechanisms for Building Resilience*. Cambridge University Press, Cambridge.

Ellis, J. E., and Swift, D. M. (1988). Stability of African pastoral ecosystems: Alternate paradigms and implications for development. *Journal of Range Management* 41, 450–459.

Gadgil, M., Berkes, F., and Folke, C. (1993). Indigenous knowledge for biodiversity conservation. *Ambio* 22, 151–156.

Hardin, G. (1968). The tragedy of the commons. *Science* 162, 1243–1248.

Lewis, H. T. (1989). Ecological and technical knowledge of fire: Aborigines versus park managers in Northern Australia. *American Anthropologist* 91, 940–961.

Little, M. A., and Leslie, P. W. (Eds.) (1999). *Turkana Herders of the Dry Savanna. Ecology and Biobehavioral Response of Nomads to an Uncertain Environment*. Oxford University Press, Oxford.

Moran, E. F. (1991). Human adaptive strategies in Amazonian blackwater ecosystems. *American Anthropologist* 93, 361–382.

Posey, D. A. (1985). Indigenous management of tropical forest ecosystems: The case of the Kayapo Indians of the Brazilian Amazon. *Agroforestry Systems* 3, 139–158.

Redford, K. H., and Mansour, J. A. (1996). *Traditional Peoples and Biodiversity Conservation in Large Tropical Landscapes*. America Verde Publications, The Nature Conservancy, Latin America and Caribbean Division, Arlington, VA.

Wells, M., and Brandon, K. (1992). *People and Parks: Linking Protected Area Management with Local Communities*. International Bank for Reconstruction and Development, Washington, DC.

Wilmsen, E. N. (1989). *Land Filled with Flies. A Political Economy of the Kalahari*. University of Chicago Press, Chicago.

HYMENOPTERA

Norman F. Johnson
The Ohio State University

GLOSSARY

arrhenotoky Reproductive mode in which unfertilized eggs develop into haploid males and fertilized eggs develop into diploid females.

eusociality Cooperative behavior among individuals of the same species characterized by reproductive division of labor, overlap of generations, and cooperative nesting.

Holometabola Insects characterized by complete metamorphosis, a wingless larval stage, and an intermediate pupal stage.

idiobiont A parasitoid that develops on a paralyzed, incapacitated host.

koinobiont A parasitoid that develops on a mobile, active host.

monophyletic A group in which all species are descended from a single common ancestor and all descendants of the ancestor are classified in the group; characterized by shared derived characters.

ovipositor Modified appendages of the seventh and eighth abdominal segments used for egg-laying paraphyletic: A group in which only some of the species descended from an ancestor are classified together; characterized by shared ancestral characters.

parasitoid An organism in which the immature stage feeds and develops on a single host arthropod, resulting in the death of the host.

parthenogenesis Reproduction in which eggs are not fertilized by males.

phylogeny Branching pattern of evolutionary relationships among organisms.

phytophagy Plant feeders, herbivores.

thelytoky Reproductive mode in which unfertilized eggs develop into diploid females.

THE INSECT ORDER HYMENOPTERA comprises a vast array of species that are familiar to even the most casual observer. The group includes the ants, bees, and wasps as well as less well-known groups such as the chalcids, ichneumons, sawflies, and wood wasps. Hymenoptera are extremely common on all continents of the world, except Antarctica. Hymenoptera range in size from microscopic species less than 1 mm in length as full-grown adults to species in which the females are 10 cm or more in total body size. Hymenoptera are holometabolous insects, generally characterized by having mandibulate mouthparts, complete metamorphosis, and two pairs of membranous wings. One striking feature is that the vast majority of species are characterized by arrhenotokous parthenogenesis: Fertilized eggs develop into diploid females and unfertilized eggs ultimately develop into haploid males. Adult Hymenoptera generally feed on nectar, honeydew, or other sugar sources, but some are actively predaceous. The larvae have a wider variety of feeding habits: Some feed on plants or fungi, whereas others feed on animals, as predators or parasitoids. The larvae of bees and a few other groups feed on provisions of pollen collected and stored by the parent female. The egg-laying appendages, the ovipositor, often are specialized to place eggs on or into the plants or arthropods that will serve as the food source for the developing larvae. In one large group, the Aculeata, the ovipositor is modified into the sting, a structure that primarily serves to deliver toxic venoms to paralyze prey or to use as a defense. The Hymenoptera are also noteworthy for the evolution of social behavior in several groups. The most highly developed form, eusociality, has arisen many times within the order among the bees, wasps, and ants.

I. CLASSIFICATION

Hymenoptera has been traditionally divided into two suborders, the Symphyta and the Apocrita. The Symphyta comprise the sawflies and wood wasps. The larvae of the vast majority of species in this suborder consume plant or fungal material, feeding on leaves or stems, feeding within galls produced in the host plant, or boring within stems or trunks of woody plants. Symbiotic fungi comprise a significant part of the diet of many of the wood-boring species. Sawfly larvae that feed externally on leaves are generally eruciform, i.e., caterpillar-like in form, with well-developed thoracic legs, abdominal prolegs, and sclerotized heads.

The suborder Apocrita is descended from a symphytan ancestral species. As adults, the first abdominal segment of the Apocrita is intimately associated with the thorax and separated from the following segments by a mobile constriction. This segment is called the propodeum. Thus, the locomotory tagma of the body, bearing the legs and wings, is four segmented. The larvae are generally immobile and highly simplified in structure. Apocrita are generally characterized by larvae that feed on other arthropods, including predatory species in which one larva feeds on several prey items and parasitoids in which the larva feeds on a single host individual. Parasitoids are distinguished from parasites in that the host of the former is usually killed as a result of feeding. Thus, ecologically, parasitoids function as predators, feeding on and ultimately killing their host. The parasitoid life history is also found in the closest living relative of Apocrita within the Symphyta—the family Orussidae. However, numerous apocritan groups are secondarily phytophagous, consuming seeds or pollen or forming galls in plant tissue within which the larvae feed and develop.

The division of Hymenoptera into two suborders is still generally accepted, even though recent evidence indicates that the Symphyta are paraphyletic. In the formal taxonomic classification, approximately 19 superfamilies and 95 families are recognized. The Tenthredinoidea, Megalodontoidea, Cephoidea, and Xyeloidea comprise the phytophagous sawflies. Wood wasps comprise four families within the Siricoidea. In the Apocrita there are many small, relatively primitive superfamilies of parasitoids: Trigonalyoidea, Ceraphronoidea, Evanioidea, Megalyroidea, and Stephanoidea. The superfamilies Chalcidoidea, Cynipoidea, Ichneumonoidea, Proctotrupoidea, and Platygastroidea encompass the vast majority of parasitoid species. Many of these are common and extremely abundant. The family Ichneumonidae is estimated to contain more than 22,000 known species, with many others yet to be recognized. The remaining four superfamilies—the Chrysidoidea, Vespoidea, Apoidea, and Sphecoidea—make up the aculeate, or stinging Hymenoptera. The Chrysidoidea include many relatively small families of parasitoids. The Vespoidea contains the ants, velvet ants, hornets, spider wasps, paper wasps, and potter wasps. The Apoidea are the bees, and the Sphecoidea are a large diverse group of predatory wasps.

The order Hymenoptera is estimated to contain approximately 115,000 described species. This is only a rough indicator of the true total because hundreds of new species are described each year in the taxonomic literature. The largest families are the Ichneumonidae (more than 20,000 species), Braconidae (approximately 12,000 species), Formicidae (the ants, with nearly 10,000 species), and the Sphecidae (more than 8000 species). Austral disjunct distributions are characteristic of a relatively few families of Hymenoptera, such as Megalyridae and Plumariidae. The major groups are found in all biogeographic realms or perhaps limited to the tropics, such as the fig wasps or Agaonidae. Ichneumonids and sawflies appear to be most diverse in the temperate realms.

II. PHYLOGENY AND FOSSIL RECORD

The oldest fossil Hymenoptera are found in deposits from the Triassic period of the Mesozoic. These include some species that are closely related to living sawflies. Parasitoid wasps, as inferred from their structure (especially the elongate ovipositor in some species) and their relationships to living species, first appear in the Jurassic. The oldest fossils of social insects, ants and bees, have been found in Cretaceous impressions and amber nodules. Minute parasitic Hymenoptera and ants and bees are commonly found in Cenozoic amber from the Baltic, the Dominican Republic, and Mexico.

Although Hymenoptera are clearly members of the monophyletic group Holometabola, comprising insects with complete metamorphosis, their position within that huge complex is unclear. Most workers place them as a basal group, sometimes as the sister group to all other holometabolous orders. Within Hymenoptera, the basal position of Symphyta is supported by evidence from morphology, behavior, molecular sequences, and the fossil record. The historical grouping of parasitoid species into a taxon variously called "Parasitica" or "Terebrantes," although of some practical use, does not reflect the pattern of phylogenetic relationships among them. The relationships among the basal Apocrita are still unclear. Most of these species are parasitoids of wood-boring Coleoptera, a life history shared with their closest living relatives in the Orussidae. Numerous hypotheses of relationships and classifications have been offered through the years that have been based on varying amounts of evidence from several sources and analyzed both more and less rigorously, but none has achieved anything approaching general acceptance. The Aculeata are a well-supported monophyletic group, and a fair amount of progress has been made in elucidating the relationships among its constituents. It is also fairly clear that the bees and sphecoid wasps are very closely related; in fact, it seems evident that bees evolved from some subgroup of sphecids.

III. BIOLOGY

A. General

1. Reproduction

Arrhenotokous parthenogenesis, or haplodiploid reproduction, is characteristic of the vast majority of species of Hymenoptera and is generally thought to have been very important in its evolution. In this mode of reproduction, offspring are produced from both fertilized and unfertilized eggs. Eggs that are not fertilized develop into haploid males. Fertilized eggs develop into diploid embryos, and these are usually females. Because of the mechanism of sex determination in Hymenoptera, which is not completely understood, diploid males are rarely produced. Mated adult females store spermatozoa received from the males during mating in a diverticulum off of the reproductive tract, the spermatheca. The female is capable of controlling the release of spermatozoa to fertilize eggs passing through the oviduct, thus controlling the sex of her offspring. The control of the sex ratio of offspring may be advantageously applied, in theory, to maximize the reproductive success of the female. In some parasitoid species, the size of the host can influence the female to lay a male or a female egg. In other circumstances, the presence of other ovipositing females on a host has been shown to influence the relative proportions of male and female eggs laid. Haplodiploidy also affects the proportions of alleles that are shared among relatives in comparison to the expectations of relatedness that are commonly found among strictly diploid species. In the simplest case of a female mating with one male, a daughter may be generally expected to share a greater proportion of its alleles with her sister (75%) than with her brother (25%). The two female offspring share one set of chromosomes contributed by the haploid male spermatozoa and, on average, would expect to share 50% of the alleles inherited from the mother. In contrast, the male offspring has only a single set of chromosomes derived from his mother. This asymmetry in relatedness has been thought to be a contributing factor to the multiple evolution of eusociality in the Hymenoptera, but its importance is strongly questioned. Although single matings may be the rule for many species under normal conditions, one female mating with many males is not uncommon: The mating flight of the honeybee, *Apis mellifera* Linnaeus, is an extreme example. Also, many species of social insects, especially among the ants, have numerous unrelated egg-laying queens within a colony.

Another form of parthenogenesis is relatively common within Hymenoptera: thelytoky. In this case, the female does not mate with a male, and her unfertilized eggs are capable of embryonic development and produce only female offspring. Thelytoky is not uncommon among parasitoid groups, and the funneling of all reproductive output into the production of new egg-laying daughters maximizes short-term reproductive success. It has recently been discovered that in several cases thelytokous parthenogenesis is caused by bacterial infections, principally by the genus *Wolbachia*. Treatment of the wasps with heat or antibiotics can cause the individuals to revert to normal arrhenotokous modes of reproduction. Many strains of *Wolbachia* do not result in thelytokous reproduction but rather in cytoplasmic incompatibility between eggs and spermatozoa. Infections by these bacteria have been discovered in a wide range of insects and the evolutionary and ecological significance of the phenomenon is not well understood.

2. Oviposition Behavior

Hymenoptera are probably most familiar as stinging insects. The sting apparatus is a modification of the egg-laying appendages fundamental to all Hymenoptera. Because egg-laying structures are naturally associated with females, only female wasps can sting. However, the males of some species are extraordinarily adept at mimicking both the structure and behavior of females and can sometimes fool even experienced workers.

In its ancestral condition, the ovipositor is composed of two pairs of elongate appendages, the gonapophyses or ovipositor valves, associated with the seventh and eighth abdominal segments. The gonapophyses are physically linked together and function as a single articulating unit. When not in use, the ovipositor is usually enclosed within a protective pair of sheaths, the gonoplacs. In sawflies, the gonoapophyses are laterally flattened, serrate, and used to cut an incision into plant material into which one or more eggs are laid. The similarity in structure and function to a saw gives rise to the common name for most of the Symphyta.

Within the Hymenoptera, the ovipositor and its function have been extended beyond its basic function in egg-laying. Among parasitoids, it is used to paralyze hosts; in the Aculeata or stinging Hymenoptera, the ovipositor and venom are often used as a defensive weapon.

B. Parasitoids

A parasitoid is defined as an animal in which the immature stage feeds on a single host individual, and this feeding activity normally results in the death of the host. Historically, such insects have been called parasites, but this was misleading. True parasites, such as lice, feed on a large host but the feeding activity has only a minor deleterious effect. Despite this distinction in names for the two type of life history, the verb "parasitize" is still normally used to describe the action of both groups. A parasitoid is essentially a predator, but one that feeds on and kills only a single prey item. They also differ from predators in that it is the parent that finds the "prey" and the offspring that feeds. The Hymenoptera are the primary group of insect parasitoids, but other orders, particularly the Diptera and Coleoptera, have species that follow this same pattern.

Parasitoid Hymenoptera attack a very broad range of arthropods. Practically all other insect groups have their suite of parasitoids as well as arachnids and myriapods. Most parasitoids attack the immature stages of their host: the egg, nymph, larva, or pupa. A few groups parasitize adult insects. In some cases, the female parasitoid oviposits on one stage and her adult offspring emerge from another. For example, some species of Braconidae oviposit within the egg of their host, but the wasp larva actually feeds on and emerges from the larva. After the completion of feeding, the parasitoid larva pupates on, within, or sometimes nearby the spent carcass of its host.

The ovipositor is used to first "sting" the host arthropod, and venom is then injected. The venom may incapacitate the host for only a short period of time or, in more extreme cases, may result in the complete suspension of most normal activities, putting the host in a state of "suspended animation." The female wasp then lays one or more eggs on or often within the body of the host. In some species, the female parasitoid attacks exposed hosts—for example, a caterpillar feeding on a leaf. In other cases, though, the hosts are hidden in the soil, enclosed within shelters made of leaves, twigs, debris, or silk, or they even bore deep within the trunks of trees. The structural characteristics, sensory capabilities, and behavioral repertoire of the female wasp combine to enable her to locate and successfully oviposit upon such protected hosts. For example, many species of parasitoid Hymenoptera that attack wood-boring beetles are equipped with elongate ovipositors that may be two or more times the length of the rest of the body. The females are capable of drilling through tens of centimeters of wood with their ovipositor in order to find and parasitize the hidden beetle larva.

The proximate mechanisms by which a female parasitoid locates, identifies, and determines the suitability hosts within a complex environment are only beginning to be understood. Chemical cues are certainly important in all of these stages of host finding. These chemicals may be produced by the hosts as a by-product of normal activity such as feeding. Often, parasitoid females are attracted to volatile chemicals produced by the plant on which a host is feeding. Other sensory modalities used in finding hosts and assessing their quality include vision, tactile examination, detection of substrate vibrations, and possibly heat detection. The external surface of the body of an adult wasp is studded with an array of sensory structures. For host location and acceptance, the most important of these are found on the antennae, tarsi, mouthparts, and ovipositor.

Among the array of details of life history strategies, two general patterns are sometimes distinguished and parasitoids may be classified as idiobionts or koinobionts. Idiobionts are generally characterized by the fact that the development of the host arthropod is arrested through the action of the venom injected by the parent. The larval parasitoid thus develops on an incapacitated host. In koinobionts, the host resumes feeding and development after the parasitoid female has oviposited. The larval koinobiont parasitoids typically develop within the body of the host, but this may be delayed for some time as the host grows and even pupates. Thus, the parasitoid must have the physiological mechanisms to evade the immune system of its host. Idiobionts typically attack hosts that are found in concealed locations and a broad taxonomic range of hosts may be attacked. Koinobionts, in contrast, may parasitize fully exposed hosts and, because they must be attuned much more closely to the host immune and endocrine systems, their host range is typically much narrower.

There is a tremendous range of variation found within the general life history patterns described. Parasitoid larvae may feed on the host from outside the host's body as ectoparasitoids, or they may live and feed within the body of the host as endoparasitoids. Among solitary parasitoids, a single individual wasp develops on each host individual. The adult wasp that emerges from the pupa then must locate mates, food, water, and new hosts on which to oviposit. Many individual parasitoids develop on a single host in gregarious parasitism; the number of wasps may range from two to thousands. In these cases, males generally emerge first and mate with the later-emerging females on or near their pupation site. In some cases, this leads to a high level of sib-mating (i.e., brothers mating sisters), and in such situations it is not uncommon for the sex ratio to be significantly skewed toward the production of females. In other words, the parent female lays just enough male eggs, sometimes only one, to ensure the insemination of all her daughters. In this manner, she is thought to be able to increase her overall reproductive output. Gregarious parasitism may be effected by numerous wasps attacking a single host, by a single wasp placing more than one egg on or in the host, or by a peculiar phenomenon called polyembryony. In this case, the parent female places one egg within the body of the host, but this single egg produces from two to thousands of identical embryos, each of which eventually develops into a new adult parasitoid. Polyembryony has apparently evolved at least four times within the Hymenoptera because it is found in the families Braconidae, Platygastridae, Dryinidae, and Encyrtidae.

Hyperparasitoids, or secondary parasitoids, are species that are actually parasitoids of other parasitoids. For example, in some species of the family Trigonalidae the adult female lays large numbers of eggs on foliage near feeding caterpillars. The eggs are inadvertently ingested by the caterpillar, the thick shell of the egg is disrupted by the caterpillar's mandibles, and the trigonalid larva quickly hatches and bores through the gut of the host to enter the hemocoel. It then lies dormant until the caterpillar is subsequently parasitized by another parasitoid species, typically an ichneumonoid or a tachinid (Diptera). The new parasitoid develops on the caterpillar host but is then attacked and ultimately killed by the trigonalid. Success in such a complicated life history requires a sequence of individually unlikely events: ingestion of the egg; survival of the host caterpillar from disease, predation, starvation, etc.; and subsequent parasitization of the caterpillar by a suitable species. The low probability of such a sequence is compensated by the fact that the female trigonalid lays huge numbers of eggs.

One of the most bizarre types of life history is the heteronomous parasitism found among some species of the chalcidoid family Aphelinidae. Female offspring develop as primary endoparasitoids of Homoptera. The males, in contrast, may develop as ectoparasitoids of the same species of host or as secondary parasitoids, sometimes attacking females of their own species.

Although the parent female of most parasitic Hymenoptera finds and oviposits on the host for her offspring, in a few cases it is the larvae that locate the hosts. In the Perilampidae and Eucharitidae, the adult female wasp lays her eggs on or in plant material, often in large numbers. The first instar larvae that hatch from these eggs are strongly sclerotized and mobile, in contrast to the typical larva of Apocrita. These mobile first instars, called planidia, somehow find their way to their host, attach, and feed. In the eucharitids, it appears that the planidia may attach to thrips, which are then picked up and carried by ants into their nest. Once there, the planidia disengage from their temporary transport and attach and feed on ant larvae.

C. Predators

The distinction between hymenopteran larvae that act as parasitoids and those that act as predators is fairly arbitrary, being dependent only on the number of host individuals killed and eaten in the process of larval development. The typical predatory life history is extremely similar to the idiobiont strategy. The parent female wasp finds, stings, and immobilizes the prey item. The prey may already be in a place of concealment, such as beneath a rock or within a burrow, or the female wasp may actually construct a hiding place, often a burrow in the soil or in tunnels in wood. More than one prey item may be cached in a chamber in this burrow, and then the female wasp lays an egg on the prey. The larva that hatches feeds on and develops on this store of food that has been provided by its mother. In the case in which only one prey item is available, this is essentially identical to the life history of an external parasitoid. In many species, the parent female may use the same burrow within which to store the prey for several of her larvae, constructing side chambers for each of the offspring or building chambers in a linear series. Thus, from the starting point of ovipositing on a paralyzed prey item within its own hiding place, there is the development of a nest—that is, a structure built by the parent female within which her young are fed and develop. Some wasps gather together all of the prey that an offspring will need for development before ovipositing. Others progressively provision their nests, providing new prey items to the developing larva as it consumes the food available.

D. Phytophagy

It is generally accepted that the Hymenoptera evolved from a phytophagous ancestor similar in many respects to the extant sawflies. The Apocrita, in turn, are derived from a parasitoid ancestor. However, within the Apocrita, there are numerous cases in which phytophagy, in its broadest sense, has reoccurred. In some cases, such as the seed chalcids of the family Eurytomidae, the host relationship is fairly straightforward: The female wasp oviposits into the developing seed and the larva feeds on the endosperm. In other cases, the phytophagy is more elaborate or even bizarre.

1. Gall Makers

Galls are abnormal growths of plant tissue caused by some sort of stimulus from another plant or animal. A wide range of insects are known to cause gall development; within the Hymenoptera it is found among sawflies, the primarily parasitic Braconidae (Ichneumonoidea) and Chalcidoidea, and especially in the gall wasps, the family Cynipidae (Cynipoidea). Gall development is induced either by chemicals injected into the plant by the ovipositing female or by secretions produced by the newly hatched larvae. The plant tissue that is elaborated into the gall may take a variety of shapes and sizes, from mere swellings of the stem to large, ornate structures on the leaves or roots. The gall maker feeds on the plant tissue within the gall, pupates there, and eventually chews an emergence hole through which it escapes. The nutritive tissue, both that of the gall and sometimes that of the developing gall maker, is also made use of by a wide variety of inquiline species which oviposit into the gall and kill the original gall maker.

The life histories of cynipid species may vary greatly. Some are fairly typical of other Hymenoptera, i.e., males and females emerge from their respective galls and then mate, and the females then seek out new host plants in which to oviposit. Their eggs develop in either females or males, depending on whether the eggs are fertilized or not. Such species are typically univoltine and attack a wide variety of host plants. Other life histories may be significantly more complicated. Some species have simply abandoned the production of males and are strictly thelytokous. Others alternate between sexual and asexual generations, reproducing via arrhenotokous and thelytokous parthenogenesis, respectively. The sexual generation consists of both males and females. These mate, and the inseminated females seek hosts in which to oviposit. All of the offspring of these wasps develop into females. The adults of this new generation in turn oviposit, and some of the eggs develop into males and some into females, thus returning to the sexual generation. The adults of the sexual and asexual generation of gall wasps often strikingly differ in their structure and produce very different types of galls in different positions on the host plant. In some cases, they even attack different species of plants.

2. Fig Wasps

The fig wasps, of the chalcidoid subfamily Agaoninae, are very specialized forms of gall makers. Female fig wasps burrow their way into the syconium, the inflorescence of the fig plant. Pollen from the body of the female is transferred to the flowers concealed within the syconium and then the female oviposits within some of the flowers. These flowers subsequently swell, and the emerging larva feeds within the tissues, consuming both the embryo and endosperm. The male wasps then emerge first and mate with the females while they are still within the galled flowers of the fig. The females later emerge from the flowers, pick up pollen, sometimes actively storing it in special cavities on the body, and escape through holes chewed by the males. Male fig wasps are extraordinarily aberrant creatures: They lack wings, the eyes and antennae are underdeveloped, but in contrast the legs and mandibles are sometimes enormous. A rich complement of other parasitic Hymenoptera are associated with figs, some developing on fig tissues and others acting as parasitoids of the fig wasps.

3. Leaf-Cutting Ants

The leaf cutters belong to the tribe Attini (Formicidae). These ants are found nearly throughout tropical

America, north into the southern portions of the United States. The long lines of thousands of ants, pruning pieces of vegetation and carrying them back to the nest are a familiar sight throughout the New World. Strictly speaking, the leaf cutters are not phytophagous because they do not feed directly on the vegetation that they collect. Rather, they use the leaves and stems as the substrate on which they maintain a fungal colony. The ants feed on specialized structures produced by their fungus; the fungal organism, in return, is nurtured and maintained in an environment favorable for growth. When leaving the nest on her mating flight, the new virgin queen takes a bit of the fungal mass to use as an inoculant when she begins her new colony. Colonies of leaf cutters can be extremely large, numbering in the millions. As a result, they can be extremely destructive to agriculture, forestry, and horticulture.

4. Pollen Feeders

The life history described previously for predatory stinging Hymenoptera has been modified in at least two cases in which the parent female provisions her nest with pollen rather than arthropod prey. The largest group in which this has occurred, and one of the groups most familiar to the casual observer, comprises the bees. Female bees visit flowers both to gather nectar for their own energetic needs and to gather the pollen on which their offspring will feed. Some bees are oligolectic, i.e., they are fairly restricted in the range of plant species from which they gather pollen. Others, such as the ubiquitous honeybee, are polylectic and gather pollen from a broad range of species. There are many specializations of the body of bees that facilitate the transport of pollen. Some bees will ingest the pollen grains and transport them within their crop; others have dense beds of hairs particularly on the legs or on the underside of the abdomen where pollen is packed for transport. Pollination of flowers by bees—that is, the transport of pollen from one flower to another—is often accidental and is irrelevant to the insect but of critical reproductive importance to the plant. Many anatomical features of plants are believed to be specializations for attracting bees or other insects to the nectar and ensuring that pollen grains adhere to their body.

E. Social Behavior

Social behavior is most simply defined as that of groups of individuals of the same species that cooperate with one another. Simple aggregations of individuals may occur, for example, where some limiting resource is found, such as water or nesting sites. Evolutionarily important social behavior, however, involves some sort of cooperation among individuals leading to reproductive success of some or all of the participants. Among the insects, the extremes in development of social behavior primarily are found in the termites (order Isoptera) and within the Hymenoptera. Its most highly developed level, called eusociality, is characterized by cooperation among females in nesting, overlap in generations, and reproductive division of labor. This means that some individuals sacrifice production of their own offspring in order to facilitate reproduction by other individuals of the same colony. Eusociality has clearly evolved several times within the order Hymenoptera: once in the ants, which are primitively eusocial; once in the large family Sphecidae; and several times among the bees and the social wasps (i.e., the paper wasps, yellowjackets, hornets, etc.).

Colonies of social insects can be extremely large, both physically and in terms on total numbers of individuals. The life cycle of a colony typically begins with the mating flight of a virgin queen. After mating with one to several males, the queen begins construction of a nest in the soil, a natural cavity, or in some cases in the open. Colony founding is sometimes cooperative (e.g., in some paper wasps), and the determination of which individual will become the primary reproductive ones is established through behavioral interactions among the founders. Males do not participate in colony founding and only serve to inseminate the new queens, after which they soon perish. The eggs produced by the queen develop into the first worker generation: These are all daughters of the queen and typically do not reproduce. Workers forage for food and nest materials, care for the developing brood, defend and care for the nest, and care for the queen. This range of behaviors is sometimes divided among the workers. In some species, individuals are morphologically specialized into recognizable castes for certain functions such as defense in the case of soldiers. Morphological differentiation into castes results from allometric growth of the individual workers. Individuals of overall larger body size have disproportionately large mandibles, spines, and head capsules. In other cases, individuals may not be morphologically distinguishable but specialize in particular sets of behaviors at different periods of their adult life, usually ending as foragers outside the nest. Once well established, the collective efforts of the individuals of the colony result in the production of a new generation of males and reproductive females. Some social species reproduce by swarming; In the case of the honeybee, the old queen leaves the existing nest, taking with her a portion of the worker force, and reestablishes in a new site. Then one of the newly emerging queens takes over the remainder of the colony.

Communication among individuals is critical to social behavior. Chemical communication is probably of primary importance. Many exocrine glands have been identified on social Hymenoptera from all major parts of the body. These chemicals include alarm and defense pheromones, mating pheromones, trail pheromones, and signals from the queen that suppress the reproduction of workers. Trophallaxis, the exchange of food between individuals, may serve as the medium of exchange of some chemical signals. Tactile behavior is also important, principally involving antennal contact between individuals. The classic example of tactile communication is the waggle dance of honeybees. The location of food sources, including direction and to some extent distance, is transmitted to nestmates by a stylized reenactment of the foraging flight within the darkness of the hive. The direction of the resource in relation to the sun is indicated by the angle formed by the waggle portion of the dance in relation to the vertical axis within the hive.

Eusocial behavior is a fairly rare phenomenon in the animal world. The repeated evolution of such complex behavior within the single order Hymenoptera has led to a great deal of work in search of an explanation. One suggestion is that the asymmetric relationship among siblings may be an important underlying feature that predisposes this group to the evolution of the characteristics of sociality. However, within the group at least two separate paths leading to eusociality have been identified. The parasocial route first envisions communal nesting among related females, followed by cooperative brood care and eventually reproductive division of labor. The subsocial route to eusociality posits first the development of overlap of generations followed by continued reproduction of the parent female in the company of her own daughters.

These characterizations of social behavior are often more simplified than the real situation found in nature. Colonies may be founded by more than one queen; sometimes these are closely related. Queens are probably usually inseminated by several or even many males, thus reducing the relevance of the asymmetry of relationships associated with arrhenotokous parthenogenesis. It is also clear that eusociality, although a "highly evolved" trait, is not the end point of evolution in Hymenoptera. Local environmental factors are undoubtedly important in the development of eusocial traits and in their subsequent loss. Several cases have been identified in which these behavioral repertoires and both the morphological and physiological features associated with them have been lost. Some of the most extreme cases are social parasitism in which reproductive females of one species live within the nest of another social species. The workers of the host species feed and care for the parasite, and the latter contributes nothing to the welfare of the colony. Social parasites are widely found among the ants, bees, and wasps.

IV. PRACTICAL IMPORT

A. Pests

In comparison to other large orders of insects, the Hymenoptera have relatively few species that qualify as pest species. A few of the phytophagous forms do cause some damage to agricultural crops or to trees. Leaf-cutting ants and leaf-cutter bees can cause damage by removing significant amounts of foliage from plants. In contrast to the ants, leaf-cutter bees do not feed on the leaves, either directly or indirectly, but use the pieces in order to line the brood chambers in their nests. Seed chalcids may destroy seeds of crops such as clover. Many of the phytophagous sawflies are important pests, particularly to forest trees. The family Diprionidae, which includes species such as the red-headed pine sawfly (*Neodiprion lecontei*), are often important pests of conifers. Other families such as the Tenthredinidae (e.g., the larch sawfly, *Pristiphora erichsonii*, and the birch leaf miner, *Fenusa pusilla*) and Cephidae (e.g., the wheat stem sawfly, *Cephus cinctus*) are occasionally important.

The most significant deleterious role played by Hymenoptera is derived from the reaction that people have to being stung. Although a few species of parasitic Hymenoptera are capable of jabbing their ovipositor through the skin of a person, this usually requires that the person actually hold the wasp in his or her hand: This is not a health concern. Aculeate Hymenoptera, on the other hand, inject venom through the sting apparatus into the subject of their attention. In some cases, the venom can cause severe pain by virtue of the properties of the chemicals. The most significant hazard, though, lies with the reaction of the human immune system to the components of the injected venom. The typical response of swelling, pain, and itching may become life threatening for individuals that become hypersensitive to the components of the venom. In extreme cases, the resulting loss in blood pressure and shock can cause death, even from the sting of a single bee.

The so-called "killer bees" or Africanized honeybees are perceived as a serious health risk in the New World. These bees are not an aggressive exotic species that has invaded the Western Hemisphere but rather the same species as the honeybee that is used in apiculture throughout the world (*Apis mellifera*). Colonies of the honeybee from Africa were brought to Brazil in order to try to transfer some of their useful traits into the colonies of European honeybees being maintained in the tropics through controlled breeding. Unfortunately, African bees escaped from confinement and became established as feral colonies. This strain of bees has spread and interbred with both feral and domesticated colonies of honeybees throughout South and Central America and into the southern United States. The African subspecies of the honeybee is notorious because the colony more vigorously defends its nest against marauders than the European bees and is generally more sensitive to disturbance. In defense, the bees attack the animal or person disturbing the colony by stinging, just as do European bees. The difference between the two types of bees lies not in a more toxic venom but in the fact the African forms are quicker to attack and are more persistent in pursuit and stinging. Numerous deaths of humans and livestock have been recorded that result from the high number of stings inflicted and not from hypersensitivity of the person or animal being attacked. In areas in which the honeybees have become more aggressive, greater care is needed by persons approaching the hives.

B. Biological Control Agents

Although some species of Hymenoptera have negative impacts on humans and their commerce, by and large the order is considered to be very beneficial. One of the most important categories of such beneficial insects consists of those that either parasitize or prey on other pest arthropods. These species, by eventually killing the other insects, act as biological agents that reduce the population levels of their hosts. Such control agents have many advantages: Parasitoids are often fairly host specific, and thus their effects are focused on the pest problem; a population of biological control agents is capable of maintaining itself through time, thus providing continuous control; parasitoids are often extremely effective at locating their hosts in a complexly structured environment; and parasitoid populations respond to those of their hosts, increasing when host populations increase and decreasing when the availability decreases. The importance of parasitoids as regulators of the populations of their hosts is most clearly seen when, e.g., a phytophagous species is accidentally introduced into a new region without its normal complement of natural enemies. Populations of such species are capable of explosive growth in the new environment, rapidly becoming important pests, e.g., gypsy moths (*Lymantria dispar*) in North America and the cassava mealybug (*Phenacoccus manihoti*) in tropical America. Biological control programs involve the discovery of the natural environment and enemies of such pests and the importation of these natural enemies for control. Some examples of biological control programs have been spectacularly successful, with effective long-term control of pests. The parasitoid families Encyrtidae and Aphelinidae have been particularly effective in biological control programs of many species of Homoptera.

C. Pollination

The successful development of seeds and fruit of many species of flowering plants depends on pollination—the movement of pollen from one flower to another. Although some species of commercially important plants, particularly grasses, are wind pollinated, a great many others rely on animal agents for the transfer of pollen. Under natural conditions pollinators can include vertebrates, such as bats and hummingbirds, and a wide variety of insects, including flies and beetles. Hymenoptera, however, and in particular the bees, are extremely important as pollinators. Recall that bees provision their nests with plant pollen on which their developing larvae feed. In the process of collecting pollen for their nests and also in collecting nectar from within the flowers, pollen grains that adhere to the hairy body of a bee can be transferred to the style of the flower. Natural pollinators include a vast array of native bees, but for many commercial crops colonies of honeybees are used to effect pollination. The honeybee is a polylectic species, that is, it gathers and its larvae feed on pollen from a wide variety of plant species. Colonies are regularly transported to fields in order to temporarily increase population levels to maximize fruit and seed production.

D. Honey and other Bee Products

The large colonies of social bees also store honey within the nests that they construct. Honey is the modified and concentrated product of the nectar gathered principally at flowers and is used as an energy source for the adult bees. The most familiar species that serve as sources of honey for human consumption belong to the genus *Apis*, including the domesticated honeybee, but both bumblebees and stingless bees also produce and store honey. Honey production and the sale of it often provide a significant income supplement for small farmers. Another product from honeybee colonies that is of some value is beeswax, used variously in cosmetics, pharmaceuticals, and candles.

See Also the Following Articles

INSECTS, OVERVIEW • PARASITOIDS • POLLINATORS, ROLE OF

Bibliography

Gauld, I. D., and Bolton, B. (1988). *The Hymenoptera*. Oxford Univ. Press, Oxford.

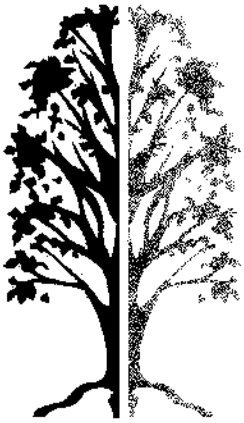
Goulet, H., and Huber, J. T. (1993). Hymenoptera of the world: An identification guide to families, Publ. No. 1894/E. Research Branch, Agriculture Canada, Ottawa.

Hanson, P. E., and Gauld, I. D. (1995). *The Hymenoptera of Costa Rica*. Oxford Univ. Press, Oxford.

Hölldobler, B., and Wilson, E. O. (1990). *The Ants*. Harvard Univ. Press, Cambridge, MA.

Naumann, I. D. (1991). Hymenoptera. In *The Insects of Australia, 2nd ed.*, (I. D. Naumann, ed.), pp. 916–1000. Cornell Univ. Press, Ithaca, NY.

Quicke, D. L. J. (1997). *Parasitic Wasps*. Chapman & Hall, London.



INBREEDING AND OUTBREEDING

Katherine Ralls,* Richard Frankham,† and Jonathan Ballou*
*Smithsonian Institution; †Macquarie University

GLOSSARY

coadapted gene complexes Chromosomes, loci, and/or genes that are adapted to function well together.

effective population size (N_e) Size of the ideal population used in population genetics theory that would have the same rate of increase in inbreeding or decrease in genetic diversity as the actual population under study.

lethal equivalent Group of mutant alleles that would cause an average of one death if homozygous; for example, one lethal equivalent might represent two mutant alleles, each with a 50% probability of causing death, or any other combination of mutant alleles that would produce an average of one death.

INBREEDING REFERS TO MATING of related individuals. Inbreeding results in a decline in survival and reproduction (reproductive fitness), known as inbreeding depression, in most species of plants and animals and can increase the extinction risk in wild populations. Outbreeding refers to matings between individuals from different populations or subspecies. Outbreeding can result in a decline in reproductive fitness known as outbreeding depression, but this is less common than inbreeding depression. This article discusses the conservation implications of inbreeding and outbreeding depression.

I. INBREEDING

Inbreeding is the mating of individuals related by ancestry. This includes self-fertilization, brother–sister, parent–offspring, and cousin matings, as well as matings between more distant relatives. Inbred offspring are more likely to inherit recent copies of the same allele from both parents, that is, alleles that are identical by descent.

Two alleles that are identical by descent are homozygous, but not all homozygous alleles are identical by descent. A homozygous individual has two alleles at a locus that are functionally similar. However, these two alleles may or may not be identical by descent. The inbreeding coefficient (usually symbolized by F) of an individual is the probability that the individual has two alleles at a locus that are identical by descent. Because F is a probability, it ranges from 0 for noninbred individuals to 1 for completely inbred individuals. For example, the inbreeding coefficient of an individual resulting from self-fertilization is 1/2 and that for an individual resulting from a parent–offspring or brother–sister mating is 1/4.

Encyclopedia of Biodiversity, Volume 3

Inbreeding reduces the frequency of heterozygotes in proportion to the inbreeding coefficient and increases the frequency of homozygotes. Natural populations contain low frequencies of deleterious recessive mutations that are normally found as heterozygotes. Inbreeding exposes them as homozygotes, so that their deleterious effects are expressed. Consequently, in most populations of animals and plants, inbreeding results in a decline in reproduction and survival (reproductive fitness), which is called inbreeding depression.

II. INBREEDING DEPRESSION

A. Evidence for Inbreeding Depression

The deleterious effects of inbreeding were known long before the discovery of the underlying Mendelian mechanisms. In the nineteenth century, Charles Darwin clearly documented inbreeding depression based on studies in 52 species of plants, as well as the experience of livestock breeders up to his time. These early observations were amply confirmed by subsequent studies. There is now extensive evidence for inbreeding depression in laboratory and domestic animals and plants, and growing evidence for inbreeding depression in wild animals and plants. Many studies on inbreeding were massive, involving large numbers of animals over many years, and the literature is extensive. For example, Sewall Wright's classic experiments on inbreeding in guinea pigs resulted in the production of 29,310 inbred and 5105 control young from 1906 to 1924. Only the five most vigorous inbred lines survived to the end of the experiments; 30 other inbred lines went extinct or declined so severely that Wright discontinued breeding them before the end of the experiment. Nevertheless, the surviving inbred guinea pigs were consistently inferior to the controls in number of young born, percentage of young born alive and raised to 33 days, and weight at 33 days.

Inbreeding and selection have been used to fix desirable traits in modern breeds of livestock. However, reduced fertility was a major problem during the early periods of inbreeding and the inbred lines were repeatedly outcrossed to restore vigor and fertility. The North Central regional dairy cattle breeding project (involving Iowa, Michigan, Minnesota, Missouri, Ohio, South Dakota, and Wisconsin), begun in 1947, is an example of the many extensive studies on inbreeding in livestock. This study found that inbreeding usually increased juvenile mortality and decreased milk yield, fat yield, growth, and reproductive performance.

Inbreeding depression is also well known in zoo animals. For example, in the early 1980s, Ralls and Ballou (1986) found that juvenile mortality was higher in inbred than in noninbred offspring in 41 of 44 populations of mammals in zoos, including many primates, antelopes, and deer, as well as a variety of smaller mammal species (Fig. 1). On average, the progeny of father–daughter and brother–sister matings suffered a 33% reduction in juvenile survival compared to outbred offspring. However, the severity of inbreeding depression varied widely across species. Although inbreeding in zoo animals usually results in less vigorous and fertile individuals that appear phenotypically normal, it is also responsible for some genetic diseases in captive populations, including blindness in wolves and dwarfism in California condors. Zoo populations are now routinely managed to avoid inbreeding.

Evidence for inbreeding depression in wild or semiwild environments has been reported in several species of fish, snails, sparrows, lions, shrews, deer mice, and many species of outbreeding plants. The question of whether or not cheetahs show inbreeding depression in the wild has been extremely controversial. Efforts to resolve the controversy have been inconclusive because all populations appear inbred and there are no outbred populations for comparison. Several studies in birds have reported a lack of inbreeding depression in the wild, but these results may be due to incorrect attribution of paternity (molecular genetic studies have shown that inferences regarding paternity based on behavioral observations are often incorrect) and biases in environmental quality that favor inbred matings. Because inbreeding depression is so common, management of an unstudied outbreeding species should be based on the assumption that it will suffer reductions in reproductive fitness if it is inbred.

B. Factors Affecting the Severity of Inbreeding Depression

The degree of inbreeding depression in a population depends on the extent of inbreeding, the original frequency of deleterious recessives, and the environment. Continued inbreeding results in greater inbreeding depression. For example, average levels of various components of reproductive fitness, such as juvenile survival in mammals or grain yield in maize (Fig. 2), theoretically increase in a linear fashion as the inbreeding coefficient increases. The more deleterious alleles that were formerly masked by heterozygosity, the more severe the effects of inbreeding will be. A locus must have some form of dominance to contribute to inbreeding depression—complete dominance, partial dominance,

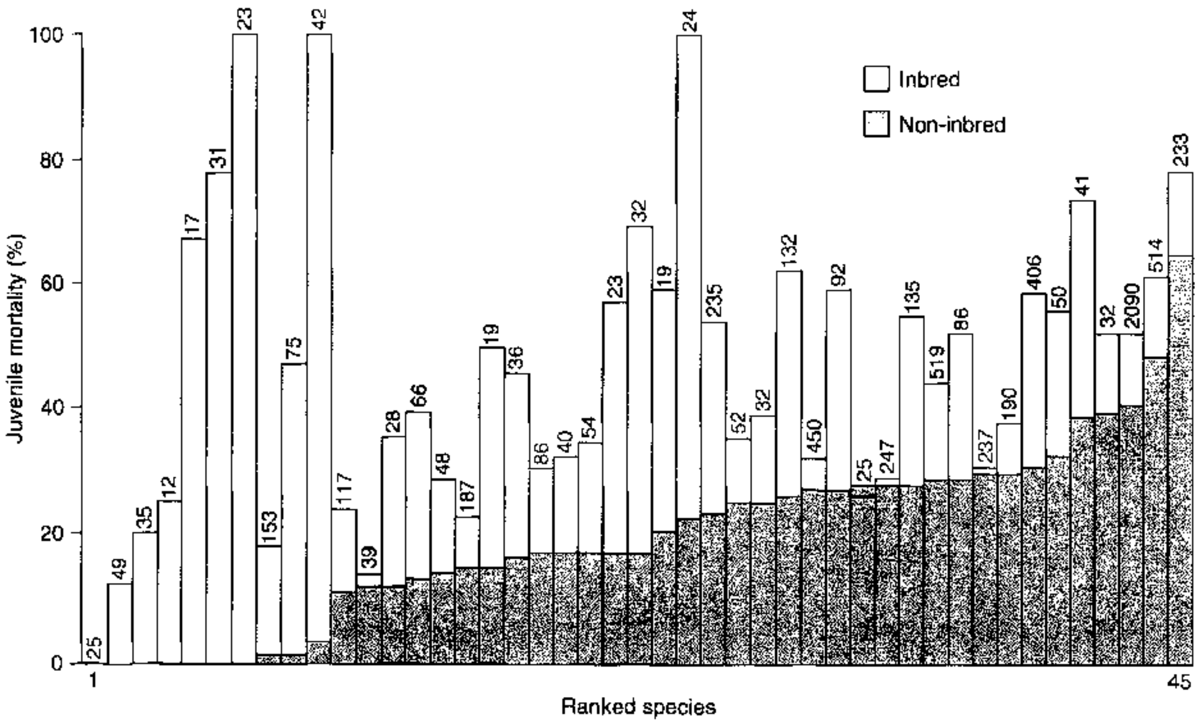


FIGURE 1 Juvenile mortality in inbred and noninbred young in 45 populations of mammals in captivity. Inbreeding levels were calculated with respect to the founders of the population for which pedigree data were available. Most populations were founded with wild-caught animals but some were founded with animals from other zoos or of unknown origin. The noninbred category includes all animals with an inbreeding coefficient of zero and the inbred category included all of those with an inbreeding coefficient greater than zero. For the larger species, all young surviving to 6 months or more were considered to have survived. For the small species, one-half the age at sexual maturity was used as the criterion age. Numbers above the bars indicate sample sizes for each species. Reprinted from Ralls and Ballou, Captive breeding programs for populations with a small number of founders, © (1986) p. 19, with permission of Elsevier Science.

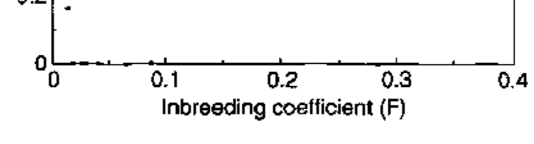


FIGURE 2 Decreasing juvenile survival with increasing inbreeding in golden lion tamarins (*Leontopithecus rosalia*). The proportion of offspring with a given inbreeding coefficient surviving to 7 days of age is plotted against the inbreeding coefficient. The sizes of the ovals represent the number of offspring available to estimate survival for each level of inbreeding. Large ovals represent samples of more than 50 offspring, medium ovals from 30 to 50, small ovals from 10 to 30, and tiny ovals from 1 to 10.

or overdominance (heterozygotes superior to both homozygotes). Most inbreeding depression is thought to be due to dominant rather than overdominant loci.

Inbreeding depression is greater in more stressful conditions. For example, Dudash (1990) found that selfed progeny of one species of plant showed 75% inbreeding depression in the field but only 59% in garden plots and 53% in a greenhouse. Because estimates of the cost of inbreeding are based on captive populations of animals and plants maintained under relatively unstressful conditions, the deleterious effects of inbreeding on natural populations are probably higher than current data suggest.

C. Measuring Inbreeding Depression as Lethal Equivalents

The extent of inbreeding depression in survival in animals can be measured in terms of lethal equivalents. The number of lethal equivalents per gamete or individual can be calculated from the rate that juvenile survival decreases with increasing inbreeding. Ralls *et al.* (1988) found an average of 4.6 lethal equivalents per individual in 40 captive populations of mammals. Thus, each individual contained deleterious mutations that would be equivalent to slightly less than 5 lethal mutations if they were homozygous. This figure is similar to estimates for humans and birds. However, the number of lethal equivalents varied widely across captive populations.

D. Inbreeding Depression in Total Reproductive Fitness

Most studies of inbreeding depression measure only one or a few components of reproductive fitness. However, all components of reproductive fitness are subject to inbreeding depression. In animals, this includes offspring survival, number of offspring per female, male mating ability, sperm quality, and the quality and quantity of maternal care. Over a wide range of species, O. H. Frankel and M. E. Soulé noted that each 10% increase in the inbreeding coefficient caused approximately a 5–10% decline in the mean values of individual components of reproductive fitness but a 25% decline in total reproductive fitness. This indicates that inbreeding depression is approximately three times greater for total reproductive fitness than for its individual components.

E. Variation in Susceptibility to Inbreeding Depression

Inbreeding depression has a large chance (stochastic) element because of the random sampling of alleles during reproduction. Individuals with the same inbreeding coefficient, that is, the same probability of carrying alleles identical by descent, differ in actual levels of homozygosity and some are more fit than others. Families and populations within a species carry different types and numbers of deleterious mutations and differ in their susceptibility to inbreeding depression. Differences in the extent of inbreeding depression among lineages within species have been reported in mice, dairy cattle, fruit flies, and flour beetles and are to be expected in all outbreeding species. These differences contribute to the occasional success in establishing inbred lines or wild populations from the progeny of a small number of founding individuals even in species that typically show high levels of inbreeding depression. Differences among populations within species have also been reported. No differences in susceptibility to inbreeding depression among major taxonomic groups are known, but the relevant data are limited.

F. Inbreeding and Outbreeding in the Wild

Only a small proportion of species regularly self-fertilize and these tend to have life histories that strongly favor mating with relatives. For example, many are colonizing species in which the chances of successfully dispersing are much greater if only a single individual has to reach new habitat. The restricted taxonomic distribution of selfing to some plants, terrestrial slugs, and marine invertebrates suggests that it is an evolutionary dead end.

The majority of species appear to be naturally outbreeding. Many plants that are pollinated by insects have elaborate morphological mechanisms that favor cross-pollination. In other plants, cross-pollination is ensured because the male and female gametes do not mature at the same time. Another mechanism that prevents self-fertilization is self-sterility, in which pollen either fails to germinate on a stigma of its own flower or germinates but does not develop sufficiently to fertilize the egg.

Many animals are also thought to avoid close inbreeding. Although a few mammals, notably the naked mole rat, normally mate with close relatives and are highly inbred, most mammals and birds rarely mate with close relatives. Sex differences in dispersal patterns often limit opportunities for inbreeding. In most mammals, males tend to disperse more frequently or farther than females, whereas the reverse is true in birds. Furthermore, many species can recognize close kin, and a variety of species such as jays, woodpeckers, mice, voles, ground squirrels, black-tailed prairie dogs, and chimpanzees are known to actively avoid mating with them.

G. Inbreeding Depression in Small Populations

Inbreeding is unavoidable in small, closed populations because all individuals eventually become related to each other. Inbreeding in a population of size N_e increases at a rate of $\frac{1}{2N}$ per generation. For example, in a population of size 10, there is a 5% increase in inbreeding per generation. Consequently, small isolated populations that have existed for many generations are expected to show inbreeding depression. Small populations of plants, fruit flies, a rock wallaby, Florida panthers, and a snake have been found to suffer from inbreeding depression. However, inbreeding depression may not cause declines in population size. Reduced fecundity and survival will only cause a population decline if the reproductive rate drops below replacement level. Small populations can also suffer from reduced reproductive fitness because of declines in environmental quality.

H. Reduction of Inbreeding Depression

Inbreeding depression may be reduced by selection against deleterious alleles, which eliminates, or purges, them from the population. Purging has been documented in plants, mice, birds, and fruit flies and during the development of inbred lines in a variety of species. Species that naturally inbreed generally show less inbreeding depression than naturally outbreeding species because of the greater opportunity for selection against deleterious recessives. Most populations of self-fertilizing plants show some level of inbreeding depression, but those with higher selfing rates tend to show less inbreeding depression (Fig. 3). Slower inbreeding generally causes less inbreeding depression than an equivalent amount of rapid inbreeding because there is more time for natural selection to operate, but this effect is often small.

Purging may reduce inbreeding depression, but it is unlikely to eliminate it. Ballou (1997) found that selection led to a small reduction of inbreeding depression in neonatal survival in 15 of 17 populations of captive mammals in zoos. No trends in purging effects were observed in survival to weaning or litter size. The purging effects were not strong enough to be of practical use in captive breeding programs. An earlier study by Templeton and Read (1984) reported that purging reduced inbreeding depression in captive Speke's gazelles. However, a reanalysis of their data has cast serious doubts on this interpretation.

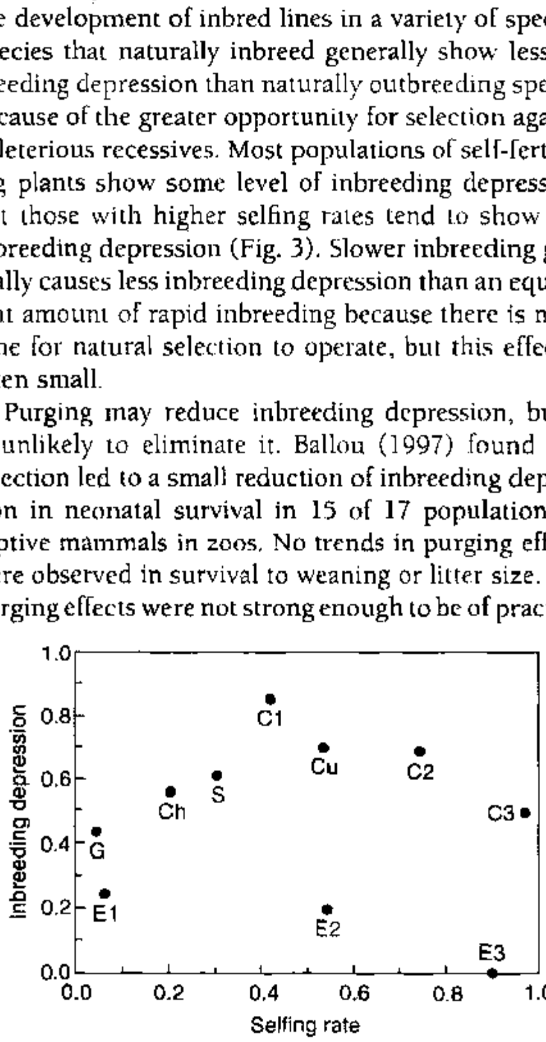


FIGURE 3 Relationship between inbreeding depression and the selfing rate in several species of herbaceous angiosperms. Species are: G. *Gila achilleifolia*; Ch. *Chamaecrista fasciculata*; S. *Sabatia angularis*; Cu. *Cucurbita foetidissima*; C. *Clarkia tembloriensis*; E. *Eichhornia paniculata*. For *Clarkia* and *Eichhornia*, data from three populations are plotted. Inbreeding depression for *these species* was measured in the lath house and glasshouse, respectively; all others were measured in the field. From Genetics and conservation of rare plants, Falk and Holsinger (eds.) © 1991 by Oxford University Press, Inc. Used by permission of Oxford University Press, Inc.

The effectiveness of selection in reducing inbreeding depression varied substantially in three subspecies of deer mice, probably due to different histories of inbreeding and selection. Lacy and Ballou (1998) found that purging was effective in a subspecies collected in relatively continuous habitat that probably experienced little or no inbreeding in the wild. Continued inbreeding reduced inbreeding depression in four of seven fitness components in this subspecies, suggesting that the original inbreeding depression was caused by highly deleterious recessive alleles that were quickly removed by selection. A second subspecies collected from ephemeral patches of fragmented habitat showed no reduction in inbreeding depression in any component of fitness. Episodes of local inbreeding in the wild may have already removed strongly deleterious recessive alleles from this subspecies. The third subspecies was a coastal form known to experience periodic population bottlenecks in the wild owing to devastating hurricanes. This subspecies had low productivity even prior to experimental inbreeding, and inbreeding in the laboratory resulted in a complete collapse of reproductive fitness. Repeated bottlenecks may have purged this subspecies of deleterious recessive alleles but also reduced heterozygosity to the point that further reductions in heterozygosity from inbreeding led to greatly reduced fitness.

Inbreeding depression can be completely reversed by outcrossing an inbred population to another unrelated population, either an outbred population or an independently inbred population. Partial recovery can be achieved by introducing unrelated individuals into the population. For example, Spielman and Frankham (1992) found that introducing a single immigrant into partially inbred populations of fruit flies increased reproductive fitness about halfway back to that found in the original outbred population.

III. INBREEDING, LOSS OF GENETIC DIVERSITY, AND EXTINCTION

Many conservation biologists have been concerned that inbreeding and loss of genetic diversity in small populations will increase the risk of extinction. However, some researchers have questioned this view because of the difficulty of obtaining direct evidence that inbreeding contributes to the extinction of wild populations.

A. Inbreeding and Extinction in Captivity

There is overwhelming evidence that deliberately inbred populations of laboratory and domestic animals and plants suffer elevated extinction rates. Most attempts to develop inbred lines by intense inbreeding are unsuccessful. For example, Bowman and Falconer (1960) found that 19 of 20 colonies of laboratory mice maintained by sibling mating became extinct by generation 12, but the remaining colony showed no decline in litter size and became a successful inbred line. Results with fruit flies, guinea pigs, poultry, and Japanese quail are similar. Extinctions occur even with extremely slow rates of inbreeding due to small population size. For example, 15 of 60 captive populations of fruit flies with an effective population size of about 50 went extinct over 210 generations.

However, extinctions of inbred lines can be caused by inbreeding, random fluctuations in birth and death rates in very small populations (demographic stochasticity), or a combination of both. Frankham (1995) found four data sets with sufficient information to distinguish genetic from nongenetic effects and examined the shape of the relationship between inbreeding level and extinction. Although the relationship between inbreeding level and inbreeding depression in individual components of reproductive fitness is theoretically linear, the relationship between inbreeding level and extinction showed a threshold effect in all cases (Fig. 4). Extinctions were rare at low levels of inbreeding but sharply increased when inbreeding reached intermediate levels. Populations inbred at different rates (full sibling versus double first cousin mating) to similar inbreeding coefficients showed similar levels of extinctions. The reason for the threshold relationship between inbreeding and extinction is that it generally takes several generations and considerable inbreeding until the growth rate of the population becomes negative. The amount of inbreeding required to reach the extinction threshold is likely to vary in different populations, as is the rate with which extinction increases with inbreeding after that point. The existence of a threshold relationship suggests that there may be little warning of impending inbreeding problems in endangered species, few of which are closely monitored.

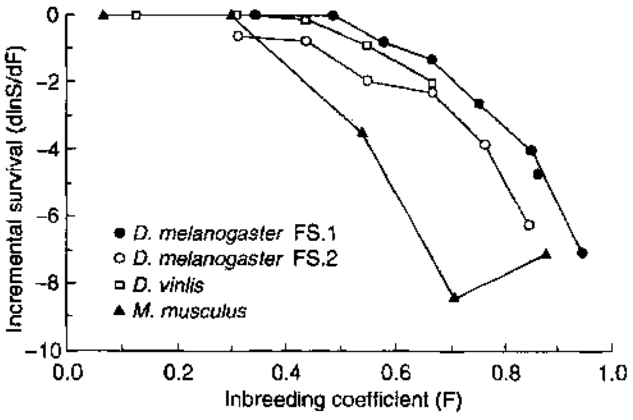


FIGURE 4 Relationships between incremental per-generation survival of populations (*d* ln *S/df*) and inbreeding (*F*) for *Drosophila melanogaster*, *D. virilis*, and *Mus musculus*. All populations were inbred using full-sibling mating, except for the use of cousin mating in the first generation in *M. musculus*. *Drosophila* data have been grouped into 0.1 increments of *F* and *M. musculus* data into 0.2 increments. Incremental values are plotted against the mean F of the two generations being considered. From Frankham (1995). Reprinted by permission of Blackwell Science, Inc.

The effects of inbreeding on population viability are complex and will interact with other factors affecting population growth, population fluctuations, or N_e, but they will be deleterious in the long term.

B. Inbreeding and Extinction Risk in Wild Populations

Conservation biologists initially thought that genetic problems played a role in the endangerment and extinction of wild populations. It was clear that inbreeding increased extinction risk in laboratory populations and theoretical work in the 1980s suggested that small populations in the wild also should suffer from increased extinction due to the unavoidable increase in mating between close relatives. However, in an influential paper, Lande (1988) argued that random demographic and environmental events will drive small wild populations to extinction before genetic factors cause problems. Environmental events, ranging from annual variation in climatic variables (such as rainfall) to catastrophes (such as fires and disease epidemics), do increase the probability of extinction and it is extremely difficult to isolate genetic effects from these other effects because inbreeding typically interacts with demography by reducing fecundity, juvenile survival, and life span. Consequently, other researchers, such as Caughley (1994), continued to question the significance of genetic factors because of the lack of direct evidence that inbreeding can contribute to the extinction of wild populations.

The first direct evidence was provided by Saccheri and colleagues (1998) with their work on the Glanville fritillary butterfly in Finland. They studied a metapopulation (a population of populations) of this butterfly that consists of numerous small populations that breed in about 1600 dry meadows of different size at varying distances from each other. Caterpillars feed in conspicuous family groups of 50 to 250 individuals and the smallest populations consist of the offspring of a single pair. Populations in individual meadows often go "extinct," but many meadows are eventually recolonized, with an average of 200 "extinctions" and 114 colonizations per year.

Because small population size results in both inbreeding and loss of genetic variation, the researchers were able to use the degree of genetic variation in each population as a measure of the extent to which it was inbred. They sampled 42 populations and found that populations with less genetic variation were more likely to go extinct. Furthermore, genetic diversity predicted extinction risk after they had accounted for all other known causes of extinction in this well-studied butterfly metapopulation. Inbreeding reduced egg hatching rate and larval survival, lengthened the duration of the pupal period (so that inbred pupae were more likely to be parasitized), and shortened female life span (so that inbred females tended to lay fewer eggs).

Furthermore, inbreeding depression likely played a role in the decline of greater prairie chickens in Illinois. Egg fertility and hatching success declined as the population became smaller. However, when prairie chickens from larger populations in neighboring states were released in Illinois, hatching success improved.

Several indirect lines of evidence also imply that inbreeding depression can be a problem in wild populations. First, theoretical studies suggest that genetic factors probably contribute to extinctions. Second, genetics may be one of the factors that makes island populations prone to extinction. These populations have lower genetic diversity than mainland populations and many are inbred to levels where captive populations show increased risk of extinction from inbreeding. Third, empirical estimates indicate that effective population size is a smaller fraction of actual population size than previously suspected, which means that genetic problems will arise at larger population sizes than previously believed. Fourth, endangered species tend to have lower genetic diversity than nonendangered species. This would not be expected if other factors drove populations to extinction before genetic factors became important. Finally, the extinction rate of a wild plant was higher in experimental populations with low genetic variation than in those with high genetic variation when both were planted in the field.

IV. OUTBREEDING AND OUTBREEDING DEPRESSION

Mating between distantly related individuals, such as individuals from different populations or subspecies, is called outbreeding. Crossing populations may increase reproductive fitness by increasing heterozygosity and thus preventing the expression of deleterious recessive alleles or may decrease fitness from the disruption of coadapted gene complexes. If outbred offspring have lower reproductive fitness than nonoutbred offspring, it is called outbreeding depression. The question of whether crossing geographically distinct populations of a species usually has beneficial or detrimental effects is important for conservation, but unfortunately there are very few data on the effects of crossing natural populations, particularly among vertebrate populations.

A. Possible Causes of Outbreeding Depression

Outbreeding depression could result from two mechanisms. The genetic mechanism requires that different populations evolve coadapted gene complexes. Crossing individuals from populations with different coadapted gene complexes could then disrupt these complexes and reduce reproductive fitness. The ecological mechanism of outbreeding depression requires that populations develop different adaptations in response to different local environments. Crossing individuals from the populations may then produce progeny that are less well suited to either local environment.

B. Evidence for Outbreeding Depression

Evidence for outbreeding depression comes primarily from organisms with extremely limited dispersal, such as some plants, copepods, and scale insects, or from crosses between individuals from vastly different geographic sources or with significant chromosomal differences. Outbreeding depression appears to be more common in plants than in animals. There is not much evidence for outbreeding depression in animals, although it has been observed in crosses where there are chromosomal differences between populations such as in dik-diks and spider monkeys. Such cases usually indicate the existence of unrecognized species or subspecies. The most widely quoted mammalian case concerns ibex in Slovakia, but this example is questionable

TABLE I

Relative Fitness of Plant and Animal Hybrid Classes[a,b]

Genus	Hybrid classification	Fitness measurement	Natural/ manipulations[c]	Fitness[d]
Plants				
Quercus	F_1	Fruit maturation	M	L (L–E)
Artemisia	Hybrid	Developmental stability	N	E
Artemisia	Hybrid	Herbivore attack	N	E
	Hybrid/F_1	Seed production and germination	N/M	E
Iris	*I. fulva*-like	Shade tolerance	M	I (I–H)
	I. hexagona-like	Shade tolerance	M	I (I–E)
Iris	Eight genotypic classes			
	Classes 1–3, 8	Viability of mature seeds	N	E
	Classes 4–7	Viability of mature seeds	N	L
Eucalyptus	*E. risdonii* backcross	Reproductive parameters	N	I (I–H)
	E. amygdalina backcross	Reproductive parameters	N	I (L–H)
	F_1-type hybrid	Reproductive parameters	N	L
Animals				
Hyla	F_1	Developmental stability	N	E
	H. cinerea backcross	Developmental stability	N	E
	H. gratiosa backcross	Developmental stability	N	E
Sceloporus	Heterozygous for chromosomes 1,3,4,6	Chromosome segregation in males	N	E
Sceloporus	Heterozygous for chromosome 2			
	Parental chromosome 2	Chromosome segregation in males	N	L
	Recombinant chromosome 2	Chromosome segregation in males	N	E
Sceloporus	Four chromosomal classes			
	HM 0	Female fecundity	N	E
	HM 1	Female fecundity	N	E
	HM 2	Female fecundity	N	E
	HM 3	Female fecundity	N	L
Colaptes	Hybrid	Clutch and brood size	N	E
Geospiza	*G. fortis/fuliginosa* F_1	Survivorship, recruitment, breeding success	N	H
	G. fortis/scandens F_1	Survivorship, recruitment, breeding success	N	H
	G. fortis × *G. fortis/fuliginosa* F_1	Survivorship, recruitment, breeding success	N	H
	G. fortis × *G. fortis/scandens* F_1	Survivorship, recruitment, breeding success	N	H
Allonemobius	Hybrid	Survivorship	N	I (L–I)
Mercenaria	*M. mercenaria* recombinant	Survivorship	N	L
	M. campechiensis recombinant	Survivorship	N	E (E–H)
Notropis	Hybrid	Survivorship	N	L (L–E)
Bombina	F_1	Viability	M	L
	Hybrid	Viability	M	E
Apis	F_1	Metabolic capacities	M	L
	First backcross generation	Metabolic capacities	M	L
Gasterosteus	Hybrid	Foraging efficiency	M	I
Gambusia	*G. holbrooki* ♀ × *G. affinis* ♂	Development	M	H
	G. affinis ♀ × *G. holbrook* ♂	Development	M	I

[a] Each of these examples involves taxa that are known to hybridize under natural conditions.

[b] Reprinted from Arnold and Hodges (1995). Copyright 1995, with permission from Elsevier Science.

[c] Natural (N) refers to those measurements taken from naturally occurring hybrids and manipulations (M) from experimental manipulations.

[d] Fitness estimates are relative to both parental species (L = lowest fitness; I = intermediate to both species; E = equivalent to both species; H = highest fitness). The most common fitness for any particular class is given, with the range of fitness values for particular classes given in brackets.

as no quantitative data were presented in the original account. Several studies have failed to find outbreeding depression in mammals, including in rhesus macaques and saddle-back tamarins. Ballou (1995) found no outbreeding depression in survival in several captive populations, including Borneo and Sumatran orangutans.

The most extensive study in mammals is that of Lacy and colleagues (unpublished), who conducted numerous crosses among five subspecies of deer mice. These subspecies included closely related populations from similar, contiguous habitats as well as very divergent subspecies that had long been isolated in dissimilar habitats. In general, the benefits of increased heterozygosity outweighed the costs of disrupted gene complexes. The reproductive success of crosses was always superior to the mean for both parental stocks, although not always above that of both parental stocks. Hybrid vigor (heterosis) was not confined to the F1 generation but continued into later generations. Experiments designed to partition the effects of heterozygosity from the effects of disrupting coadapted gene complexes indicated negative effects of disrupting coadapted gene complexes on some components of reproductive success in one long-isolated subspecies, but these effects were smaller than the beneficial effect of increased heterozygosity. A cross between two other subspecies indicated that disruption of gene complexes in the F2 and backcross generations actually increased litter viability.

C. Variation in Susceptibility to Outbreeding Depression

In general, the probability of outbreeding depression increases as populations become more distantly related. Small, isolated populations of species with poor dispersal abilities and naturally high rates of inbreeding that are subject to different selective pressures are more likely to suffer from outbreeding depression when crossed than are large populations of wide-ranging species with good dispersal abilities and low rates of inbreeding.

Small populations of naturally inbreeding species must be distinguished from populations of normally outbreeding species that have recently been reduced in size, isolated, and inbred as a result of human activities such as habitat destruction. In recently isolated populations, genetic differentiation may be the result of random drift rather than local adaptations. For example, genetic differences among extant gray wolf populations in North America are thought to reflect recent population declines and habitat fragmentation rather than a long history of genetic isolation. In such cases, outbreeding is likely to benefit, rather than decrease, fitness by restoring heterozygosity and masking deleterious alleles. For example, the small, isolated population of endangered Florida panthers shows signs of inbreeding depression, and individuals from its nearest subspecies in Texas have been added to the population in attempt to increase its reproductive fitness.

Some authors, such as W. M. Shields, have suggested that outbreeding depression is widespread and might be as important as inbreeding depression as a conservation concern. However, a recent review by Arnold and Hodges (1995) found that even hybrids between species are not uniformly unfit but may have lower, equivalent, or higher levels of fitness than their parents. Hybrid genotypes showed a wide range of fitness values, but the general pattern was that hybrids demonstrated either equivalent fitness to the two parental taxa or higher levels of fitness than at least one of the parents (Table I). The results of mixing populations or subspecies always depend on the specific animal or plant groups involved. However, outbreeding depression is rarer than inbreeding depression and is of less conservation concern.

See Also the Following Articles

ECOLOGICAL GENETICS • GENETIC DIVERSITY • POPULATION GENETICS

Bibliography

Arnold, M. L., and Hodges, S. A. (1995). Are natural hybrids fit or unfit relative to their parents? *Trends in Ecol. and Evol.* 10(2), 67–71.

Falconer, D. S., and McKay T. F. C. (1996). *Introduction to Quantitative Genetics*, 4th ed. Longman, Harlow. United Kingdom.

Frankham, R. (1995). Conservation genetics. *Annu. Rev. Genetics* 29, 305–327.

Ralls, K., and Ballou, J. (1986). Captive breeding programs for populations with a small number of founders. *Trends in Ecol. and Evol* 1(1).

Wright, S. (1977). *Evolution and the Genetics of Populations*. Vol. 3. University of Chicago Press, Chicago.

INDICATOR SPECIES

John H. Lawton* and Kevin J. Gaston†
*Imperial College, United Kingdom and †University of Sheffield, United Kingdom

GLOSSARY

acid deposition Anthropogenic acidification of terrestrial and freshwater ecosystems by (primarily) sulfuric acid, derived from sulfur dioxide produced by burning oil and coal and deposited in rain and snow (acid rain), directly as particles (dry deposition) and as cloud droplets.

all taxa biodiversity inventory (ATBI) The idea, first suggested by D. H. Janzen, that it might be feasible to produce a complete species list for all the organisms living in one place, a hectare of tropical forest, for example. The goal has so far proved elusive.

bioassay The use of living cells or organisms to make quantitative or qualitative measurements of the amounts or activity of substances.

community An assemblage of species populations that occur together in space and time.

ecotoxicology The use of test organisms (e.g., the water flea, *Daphnia*) to study the toxicity, pathways of accumulation, and breakdown of chemicals, particularly those manufactured by humans (e.g., pesticides).

endemic species Species confined in their distribution to a particular geographic region. The size of the region is arbitrary (a species can be endemic to North America or to a tiny island).

hot spots A word with several distinct meanings. Here it is used to denote sites unusually rich in a particular group of species (e.g., birds), compared with average sites in the same geographic region. (The converse is a "cold spot.") Has also been used to denote centers of endemism (see *endemic species*), which need not be unusually species rich; it is not used in this context here.

paleoclimatology The study of past climates from fossils and other traces left in the geological record.

reserve selection algorithms Mathematical techniques used to maximize efficiency in the selection of protected areas for conservation. The efficiency criteria vary with circumstances but may, for example, be the minimum number of reserves with every species represented, or minimum cost.

THIS ARTICLE REVIEWS THE USE of species as indicators of the state of the environment and of human-induced changes to the environment. We focus on five interrelated topics, namely various types of pollution, rising concentrations of atmospheric carbon dioxide, global climate change, patterns in regional and global biodiversity, and the designation of protected areas for nature conservation. Using organisms to indicate the state of, and changes to, the environment has numerous tried and tested applications. Attempts to identify indi-

Encyclopedia of Biodiversity, Volume 3

cator species to predict the diversity of other, unstudied taxa, for scientific or conservation reasons, has proved to be more contentious and much more difficult.

I. SPECIES AS INDICATORS OF THE STATE OF THE ENVIRONMENT

There are three distinct uses of the term "indicator species" in research in ecology and biodiversity. They are a species, or group of species, that do the following:

1. Reflect the biotic or abiotic state of an environment
2. Reveal evidence for, or the impacts of, environmental change
3. Indicate the diversity of other species, taxa, or entire communities within an area

This article explains, provides examples of, and evaluates each of these uses of the term, focusing primarily on terrestrial and freshwater ecosystems; broadly similar conclusions apply to marine ecosystems, but marine examples lie beyond the scope of the article. We pay most attention to the third use of the term "indicator species," because this seems most appropriate for an encyclopedia devoted to biodiversity. The most up-to-date evaluation and review of indicator species in the scientific literature is by McGeoch (1998). She concentrates on terrestrial insects as "bioindicators" (in all three senses of the word) but the general principles that she discusses extend to all ecosystems and organisms, not just to terrestrial insects.

Everybody knows that living organisms are sensitive to the state of their environment. Pollution from human activities kills many species and reduces the abundance of others. These changes in abundance can be used to assay the state of the environment.

A. An Example: Acid Deposition

Sulfur dioxide, produced by burning fossil fuel, particularly coal, enters the atmosphere and is eventually deposited on terrestrial and freshwater ecosystems via three routes: (a) as tiny solid particles, (b) washed from the air in rain or snow, or (c) as droplets formed in clouds. Deposition often occurs hundreds of kilometers from the source. Dissolved in water, sulfur dioxide forms sulfuric acid, resulting in what is frequently referred to as "acid rain," but because there are three principal routes involved in its transfer to terrestrial and freshwater ecosystems, it is more correctly called "acid deposition" (Erisman and Draaijers, 1995). Sulfur dioxide is not the only source of acidification; oxides of nitrogen, again produced by burning fossil fuel, are also involved, but sulfur dioxide is the main agent of acidification in most ecosystems.

In terrestrial ecosystems, this deposition kills lichens and acidifies the soil, leading to changes in the vegetation. Lakes become progressively more acidic as deposition loads increase, until eventually they may become virtually lifeless. A trained biologist, visiting for the first time an area subject to acid deposition, will often be able to deduce that the habitat is being polluted simply by looking at the species that are present and those that ought to be there but are not. Beautifully clear Scandinavian lakes, lacking any fish or amphibians, supporting few birds and a species-poor and taxonomically unusual invertebrate fauna, have been reduced to this impoverished state by the transnational export of sulfur dioxide from coal-burning power-stations in the United Kingdom and elsewhere in Europe. Here, living organisms act as powerful indicators of the state of the environment and the damage being done to it by human activities, often performed many hundreds of kilometers away.

B. Management of European Rivers

Because the species composition and richness of biological communities change as the environment changes, we can use species as indicators of the state of the environment for practical management purposes. The techniques have been particularly well developed to assess organic and inorganic pollution in European rivers, managed for recreation, fisheries, and drinking water. The advantages of using living organisms as indicators of water quality are that they avoid the need for expensive chemical analyses, and, probably more important, organisms integrate the impacts of pollutants over space and time. All chemical traces of a major pollution incident may disappear from a river in a matter of hours as the pollution is flushed from the system. Nonetheless, the biotic community may show evidence of the damage for many months. It is extremely difficult, and prohibitively expensive, for chemists to measure all the organic and inorganic chemical pollutants entering a river, and it is certainly impossible for them to work out what all the combined impacts of such a cocktail might be. But living communities reflect the integrated effects of all the compounds that find their way deliberately and accidentally into watercourses, and hence they act as sensitive indicators of the state of the environment. A valuable source of further information on the use of living organisms to monitor the environmental health of rivers and lakes is provided by Rosenberg and Resh (1993).

Two widely used European examples of this approach are the German Saprobic Index, and RIVPACS in the United Kingdom. RIVPACS is now used by the Environment Agency to manage UK rivers. Both the German and UK approaches require accurately identified samples to be taken of the organisms found along sections of the river. RIVPACS focuses on invertebrates, the German index on invertebrates, microbes, and higher plants. Both rely on the fact that some species are extremely tolerant of pollution (the aquatic larvae of some chironomid midges, for instance), while others are extremely sensitive, particularly to the low oxygen levels produced by organic pollution (for instance, the larvae of many mayflies). The species present in the samples are given scores, depending on their known tolerances, and the data from all species are combined to produce a composite and very sensitive index of pollution levels for any particular section of a river.

C. Widespread Application

Use of organisms to indicate the state of the environment is widespread, taxonomically and geographically, for a wide range of environmental issues. Use of species as indicators of the state of the environment is not confined to freshwater, or to Europe and North America. A wide variety of organisms has been suggested, or used, as indicators of human impacts. In Europe, suites of plant, fungal, and insect species are only found in, and hence are good indicators of, ancient woodland; they are entirely absent from plantations, even though these may be several hundred years old. The use of lichens as sensitive indicators of air pollution is well known, but organisms as different as mites and geckos (agile, climbing lizards) have been used, or suggested, for similar purposes. Lichens have also been used as indicators of fire history in Brazilian cerado (a type of dry, scrubby forest), tiger beetles as indicators of tropical forest degradation in Venezuela, and day-flying Lepidoptera (butterflies and moths) as indicators of the state of seminatural grasslands for conservation in Europe. Many other similar examples exist.

D. Interpretation Requires Care

In these and similar cases, considerable care is needed before a species or group of species can be used as reliable indicators of damaging (or beneficial) human impacts on ecosystems. All populations of living organisms fluctuate over time and vary in abundance spatially, because of natural variations in the weather, normal changes in the physical environment, and fluctuations in the abundances of natural enemies, competitors, and essential resources (food and shelter). Just because one or more species is declining does not mean that human impacts are to blame. In the case of lichens and atmospheric pollution or freshwater invertebrates and river quality, the links between anthropogenic pollutants and changes in the distributions and abundances of organisms are thoroughly researched and well understood. But even quite major declines in some species have proved exceptionally difficult to link to damage to the environment caused by people.

E. Amphibian Decline

The so-called amphibian decline is a particularly dramatic example (Blaustein and Wake, 1995). In many parts of the world, population biologists interested in amphibians (frogs, toads, newts, salamanders, etc.) have recently become alarmed by apparent major declines in the abundance, and the complete disappearance, of many species from areas where formerly they were common, often in regions apparently remote from human impacts. The declines are not happening everywhere, and the magnitude of many of those that have been claimed is difficult to assess because of the lack of long-term data prior to the supposed population collapses; some of them may be perfectly natural. The worrying aspects of the phenomenon are that while it is apparently global in scope, the causal mechanism (or mechanisms) remains obscure. It has been suggested, for example, that the amphibian decline is indicative of rising global levels of damaging ultraviolet light (UV-B) caused by loss of the earth's protective stratospheric ozone layer. Amphibian eggs, exposed in shallow water, and the adults with their thin wet skins may be particularly sensitive to UV-B, as are human sunbathers without sunblock. Others doubt the explanation. More recently a global pandemic has been implicated. But what should suddenly trigger lethal outbreaks of disease in amphibians is unclear.

F. Environmental Toxicology

In all the examples so far, the organisms being used as actual or possible indicators of environmental health have been in their natural environment. There is another related but quite separate way in which biologists use the sensitivity of organisms to set environmental standards, namely in the science of environmental toxicology, or ecotoxicology for short. In many areas of human endeavor, the aim is to apply some beneficial technology with minimum environmental damage. Crop spraying with pesticides is a good example, and so is the discharge of treated effluent from a factory. Some environmentalists claim that these types of operations should not lead to any environmental contamination; factories should have zero discharges, and if we must use pesticides, they should be targeted to reach only the crop and the pest and not, for example, the soil, nontarget organisms, or adjacent watercourses.

However, zero discharges or precision pesticides, if they can be achieved at all, can often only be obtained at great economic cost. The more pragmatic solution is to ask whether there are minimal levels of discharge, spray drift into watercourses, and so forth that cause no detectable environmental damage. To provide answers to this admittedly difficult question, environmental toxicologists use a wide variety of laboratory bioassays with standard organisms. Examples from freshwater include the alga, *Chlorella vulgaris*, the water flea, *Daphnia magna*, the amphipod shrimp, *Gammarus pulex*, and the rainbow trout, *Salmo gairdneri*. The fundamental problem is to try and establish acceptable levels of contamination. Defining "acceptable" obviously requires political as well as biological judgment. However, traces of a compound in water, air, or soil that cause no detectable changes in the performance (growth, survival, or reproduction) of the test organisms are clearly more acceptable than doses that kill 50% of the population (so called LD_{50} levels). Basically, the bioassays seek to set environmental standards for levels of potential pollutants in soil, air, and freshwater, using a range of standard laboratory organisms as indicators (Shaw and Chadwick, 1998), but there can be no absolute standards about what is safe or acceptable. The general trend in modern societies is for standards to gradually tighten.

II. SPECIES AS INDICATORS OF ENVIRONMENTAL CHANGE

If the amphibian decline (discussed in the previous section) is real, it is an example of a group of organisms acting as indicators not only of the state of the environment, but also as indicators of ongoing changes to the global environment, albeit of an unknown nature. In other words, given that species are sensitive to the condition of their environment, monitoring organisms not only tells you about the current state of an environment, but *repeated* monitoring can tell you about *changes* in that environment. To act as indicators of change rather than current environmental health, it is necessary to have at least two sets of data on the particular indicator species in question, taken in the same way, at the same place(s), on two separate occasions. More frequent sampling allows greater confidence in the direction of apparent trends and the detection of more subtle environmental changes.

A. Not All Monitoring Is about Environmental Degradation

Not all monitoring of species seeks to record environmental degradation. Increasingly after mining operations, for example, mine operators are required to restore spoil heaps and mine pits by sowing or planting native vegetation. Monitoring selected groups of common animals on nearby undisturbed control sites and on the restored land can give a good indication of the recovery of the entire ecosystem and of the success of the restoration project. For instance, when biologists monitored ant assemblages on abandoned, replanted bauxite mines in Australia, they found that the ants provided a good indication of the recovery of these ecosystems. Even after 14 years there were still differences between the ant communities found in the natural *Eucalyptus* forest and the restored land.

B. Historical Records of Change

1. Lake Acidification

It may not always be necessary to sample in real time. When anthropogenic acidification of lakes was first discovered, many people doubted that the phenomenon was real. In particular, there was considerable opposition to the notion from the power-generating industry, because solving the problem (by burning low-sulfur coal, adding "scrubbers" to power station chimneys to remove sulfur dioxide, or switching to natural gas) was inevitably going to be expensive. After all, there were few historic data on the state of the acidified lakes. Perhaps they had always been that way?

Resolving the problem required knowledge of the fact that lake phytoplankton (the tiny, unicellular plants that float in the upper layers of lakes) are extremely sensitive environmental indicators, because different species grow best in very different conditions determined by nutrient status and pH (acidity). When algae die, they sink to the bottom where their bodies and characteristic pigments are buried and some are preserved (incipient fossils), particularly the resistant, silicious outer cases of a group called diatoms. An undisturbed core through the sediments records the history of a lake's phytoplankton, with the oldest flora at the bottom. Cores showed unequivocally that many Scandinavian lakes that are acid now were not acid before the Industrial Revolution; the oldest diatoms—species not found in acid lakes—are gradually replaced in the sample column by acid-tolerant species. Diatoms are wonderfully sensitive indicators of environmental change (Fig. 1).

2. Plants and Carbon Dioxide

Herbarium specimens (pressed plants collected for taxonomic purposes) and fossil leaves can also be used as indicators of past environmental change. Another consequence of the rapid rise in the burning of fossil fuel since the Industrial Revolution has been an accelerating rise in the concentration of atmospheric carbon dioxide, one of the main agents of "global warming." We will deal with species as indicators of anthropogenic global climate change (as it is more accurately known) later. Here we want to focus on physiological and developmental responses within single species to rising carbon dioxide.

If plants are grown in a greenhouse under different atmospheric carbon dioxide concentrations, from below the pre–Industrial Revolution levels of about 280 parts per million by volume (ppm), through what are roughly present levels of 350 ppm, to levels that may be reached by the end of the 21st century (700 ppm), several interesting things happen. In particular, in the present context, stomatal densities on the undersides of the leaves decline. Stomata are the tiny pores in the leaf surface through which plants take up carbon dioxide (needed for photosynthesis), and through which they lose water vapor. It has been known for a long time that plants control the opening and closing of stomata to optimize carbon dioxide uptake and reduce water loss. More surprising, we now also know that plants grown in high carbon dioxide have lower densities of stomata; something happens during leaf development to reduce the number of stomata. How and what is currently unclear. Why is simple enough. In a high carbon dioxide world, the plant needs fewer stomata to take up the carbon dioxide it requires and hence can satisfy the needs of photosynthesis and reduce water loss by developing fewer pores in the leaves.

Now back to those herbarium specimens and fossil leaves. If you look at 200-year-old (and very precious) herbarium and modern specimens of the same species,

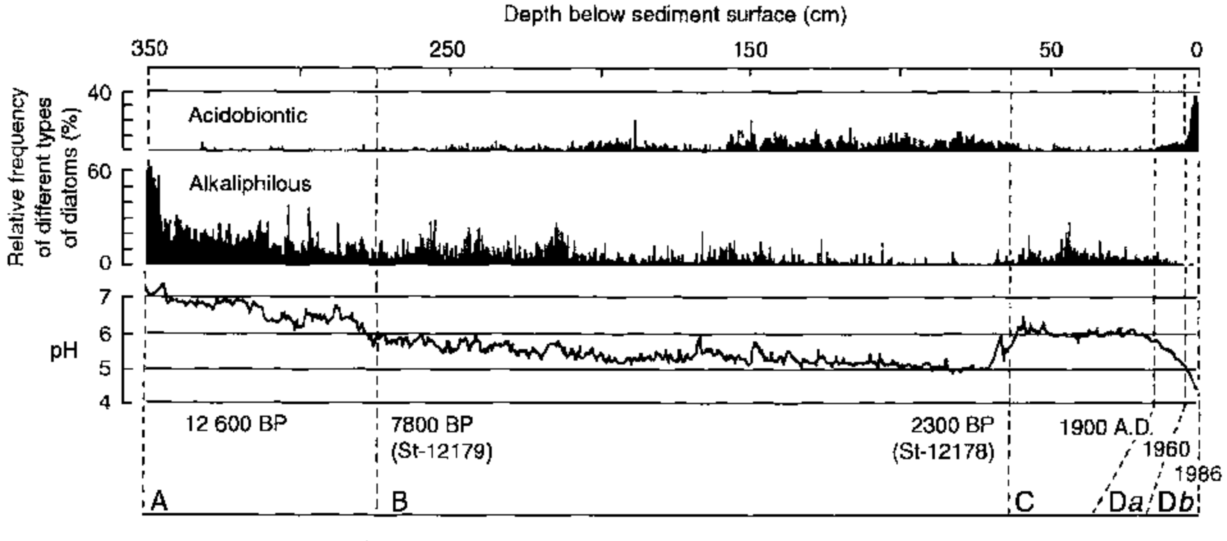


FIGURE 1 The pH history of Lilla Öresjön, a 0.6 km² lake in southwest Sweden. A core of the bottom sediments 3.5 m long records the history of the lake extending back to 12600 BP, using the valves ("shells") of diatoms preserved in the deposits. Different species of diatoms have different pH preferences and can be classified accordingly. Acidobiontic species thrive in acid waters; alkilophilous species prefer more alkaline conditions. Combining data from the remains of all species of diatoms allows the pH history of the lake to be reconstructed. The lake has passed through four pH periods. A, an alkaline period after deglaciation. B, a naturally more acidic period. C, a period with higher pH, which started at the same time as agricultural expansion in the region, and D, a rapid, recent period of acidification. The post-1960 phase has no similarity with any of the previous periods. From Renberg, I. (1990), *Phil Trans. R. Soc. London B* 327, 357–361.

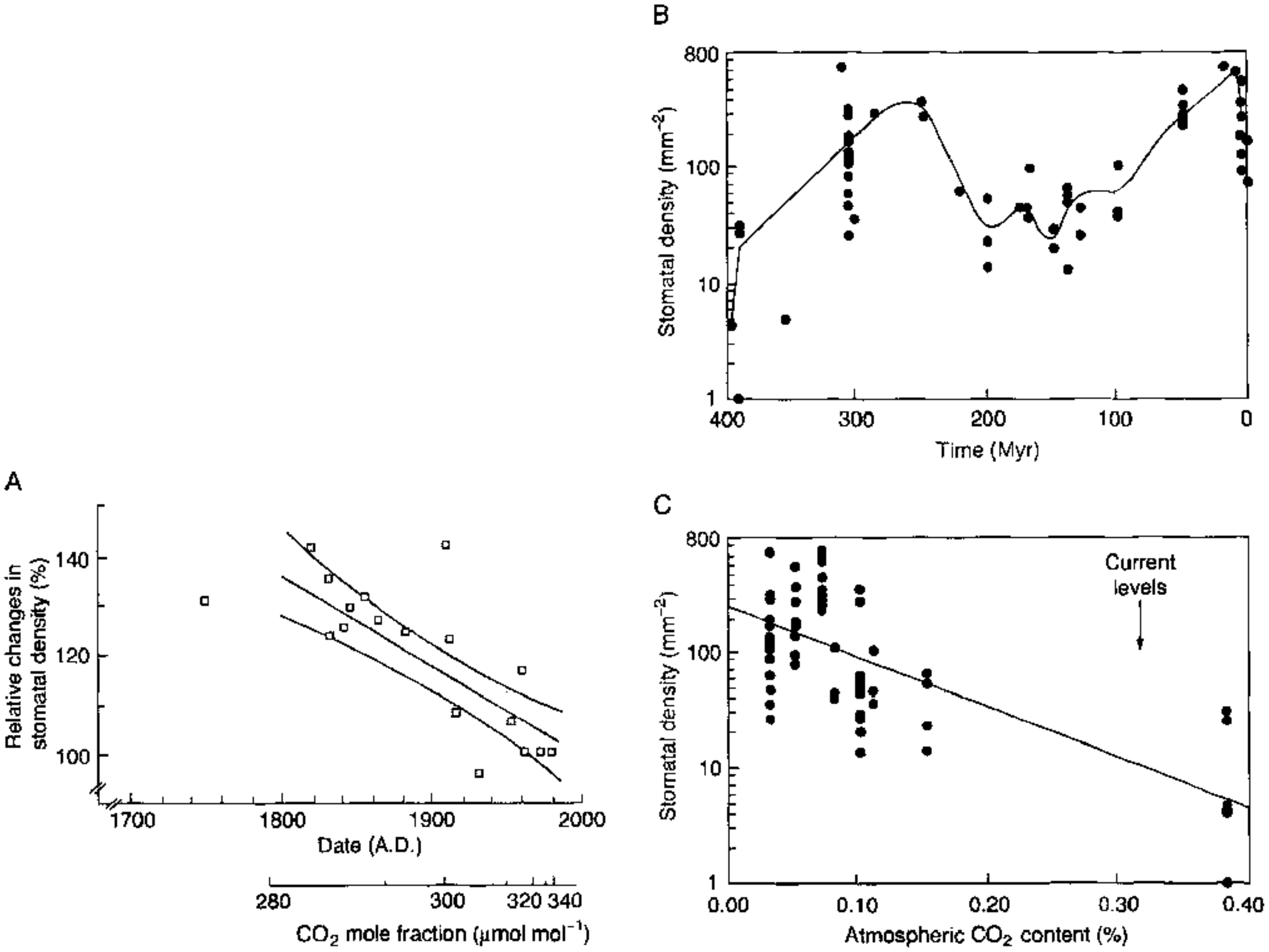


FIGURE 2 Changes in stomatal densities of leaves as indicators of changes in atmospheric carbon dioxide concentrations. (A) Percentage changes in stomatal densities on the lower surfaces of the leaves from eight species of trees and shrubs collected from the English midlands and preserved as herbarium specimens. The oldest specimen was collected in 1750. As atmospheric CO_2 has risen during and since the industrial revolution, so stomatal densities have fallen. From Woodward, F. I. (1987), *Nature* 327, 617–618. (B, C) Assuming that similar effects occur in all species of plants, fossil leaves can be used as indicators of atmospheric CO_2 concentrations extending back many millions of years. The CO_2 content of the earth's atmosphere appears to have fluctuated markedly and apparently naturally during the past 400 million years. From Beerling, D. J., and Woodward, F. I. (1997), *Bot. J. Linn. Soc.* 124, 137–153.

sure enough, stomatal densities decline as global atmospheric carbon dioxide levels increase (Fig. 2A). The same approach has recently been used to try and trace atmospheric carbon dioxide levels throughout most of the Phanerozoic, from the time when plants first colonized the land. Here the method is more contentious, because different species of truly fossil plants with presumed similar growth forms have to be used in different geological periods. Nevertheless, the pattern of apparent changes in global atmospheric carbon dioxide concentrations over hundreds of millions of years, revealed by this method (Fig. 2B and C), are in reasonable agreement with alternative, independent, and also contentious geochemical methods. Here is a really unusual use of species as indicators of environmental change.

C. Species as Indicators of Climate Change

1. The Sensitivity of Species to Climate: Fossils Again

Current, rapidly rising concentrations of atmospheric carbon dioxide are the primary cause of anthropogenic global climate change. However, the earth's climate has always changed, naturally, with no intervention from human beings. One of the ways we know this is through the careful documentation of the types and distributions of organisms in the fossil and subfossil record. The science of paleoclimatology, which seeks to reconstruct the history of earth's climate, relies heavily on changes in fossil and subfossil species assemblages to deduce what the earth's climate was like thousands or even millions of years ago. To take one example, in the modern world many types of corals occur exclusively in tropical marine environments; it is a winning bet that fossil corals of the same type indicate an ancient tropical sea, even though the rocks bearing the fossils may now lie in much colder parts of the world.

In more recent geological time, we can use changes in the distributions and abundances of plants and animals to trace major changes in the earth's climate during the Holocene (the most recent geological past) and Pleistocene glacials and interglacials. Plant remains preserved in packrat middens in dry air of the southeastern United States attest a much wetter climate only a few thousand years ago. Hippopotamus bones and teeth dug up from under Trafalgar Square provide unequivocal evidence of a much warmer London. Pollen grains preserved in peats and lake sediments record in exquisite detail the march northward of European and North American forests from the end of the last glaciation 12,000 years ago (Huntley and Birks, 1983). The forests spread with remarkable speed (an average of about 200 m per year, but sometimes as fast as 2 km a year) to achieve present distributions in the northern parts of both continents from glacial refugia thousands of kilometres to the south (Fig. 3A and B). The information is not won easily. It requires huge patience and great skill to identify thousands upon thousands of pollen grains extracted onto microscope slides. But once done, the record reads like a speeded-up movie, as spruce, oaks, white pine, hemlock, beech, and chestnut swept north in successive waves through what is now the United States and Canada; in the more species-poor forests of Europe, pines were followed by birch, then oak. These invasions are as dramatic as any in human history, but they were silent and recorded only by pollen grains.

2. Contemporary Changes in Species Distributions

Historical changes aside, there is now no doubt that the world is currently warming quite rapidly. An upward trend in global annual mean surface temperatures is apparent from about 1920, particularly over the last two decades (from c. 1980); global mean surface temperatures in July 1998 were the highest ever recorded. Do organisms act as indicators of these changes, perhaps, as with the freshwater species discussed earlier, acting subtly to integrate several of the changes human's find difficult to comprehend in the bald statistics? Climate change does not simply involve warming; it involves changes in rainfall, extreme weather events (droughts and storms), and even locally cooler conditions. All these complex changes should show up in changes in the distributions and abundances of organisms.

They do. Species are proving to be extremely sensitive indicators of contemporary climate change, where historical records allow decent reconstruction of former and current distributions. Populations of Edith's checkerspot butterfly *Euphydryas editha* are disappearing from southern California and northern Mexico, at the current southern end of its distribution, and from more lowland sites; sites where previously recorded populations still exist are on average 2° further north than sites where populations went extinct (Fig. 4). These are exactly the changes we would expect in a warming world. Twenty years ago in northwest Europe, little egrets *Egretta garzetta* (small white herons) used to be rare visitors from the Mediterranean. Now they are breeding in northern France and southern England in an astonishing expansion of range. Populations of many other European birds, butterflies, and other organisms are spreading north at the present time, as the climate warms.

Of course, none of this tells us whether the climate change that is certainly happening is "natural"—it could have happened anyway and may have nothing to do with anthropogenically produced greenhouse gasses—or whether it is indeed due to human activities. Using species as indicators of climate change tells us unequivocally that the earth's climate is changing, but so does the mercury in the thermometer. What neither tells us is why, and no end of work on species as indicators will solve that dilemma. As we have already seen, this situation is not unique to climate change. It generally holds whenever we use species as indicators of the state of the environment. Indicator species can tell us whether an environment is, or is not, changing. They do not tell us why the changes are taking place. That almost always requires additional detective work, although knowledge of an organism's biology will frequently provide valuable clues. Three examples, using birds as indicators, illustrate the problem in more detail.

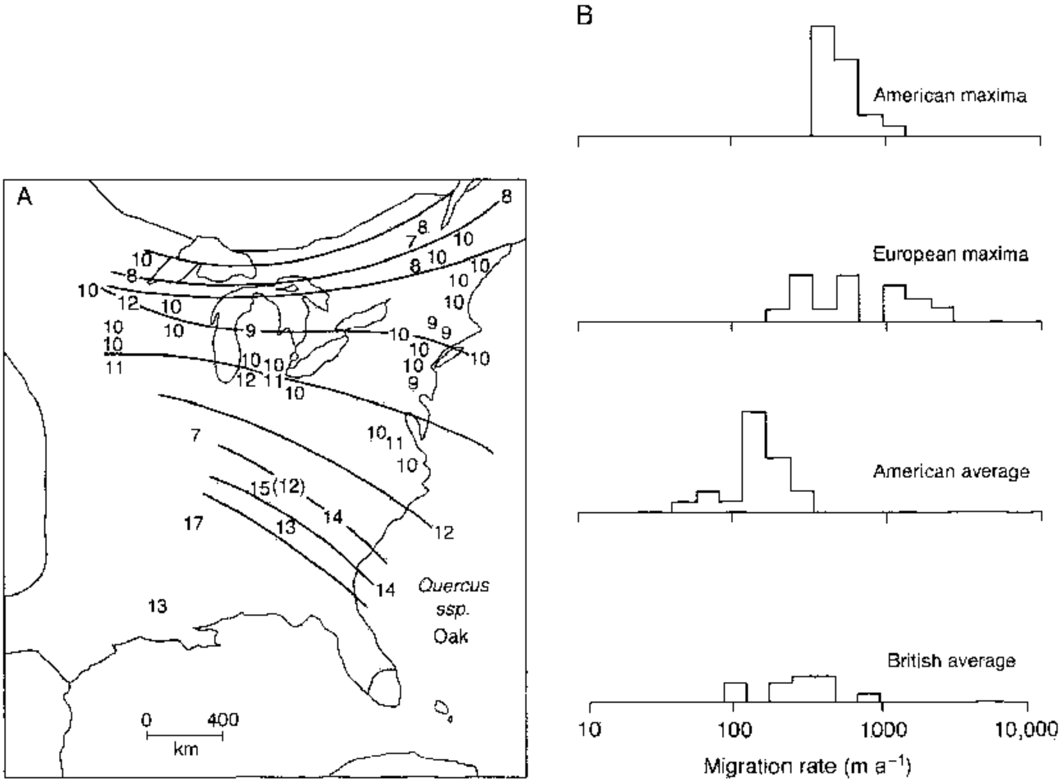


FIGURE 3 The migration of trees across North America and Europe after the end of the last glaciation, revealed by pollen remains in lake sediments and peat. (A) The spread of oaks in North America, with radiocarbon ages in thousands of years (contours) and the present range of the genus (shaded). From Davis, M. B. (1981) in *Forest Succession: Concepts and Applications* (D. C. West, H. H. Shugart, and D. B. Botkin, Eds.), Springer-Verlag, New York. (B) Estimates of the oversall rates of spread of trees on two continents, based on data of the type shown in part A. From Williamson, M. (1996), *Biological Invasions*, Chapman & Hall, London.

D. Birds as Indicators of Large-Scale Environmental Changes

Birds are widely used indicators, because in Europe, North America, and other parts of the world where there are large armies of amateur bird watchers their populations and distributions have been recorded well enough, for long enough, to reveal major environmental trends.

1. Peregrine Falcons and DDT

The catastrophic collapse of peregrine falcon *Falco peregrinus* populations throughout the northern hemisphere in the 1950s signaled widespread contamination of the environment by chlorinated hydrocarbon insecticides, first DDT, then other compounds such as aldrin and dieldrin. The total, and rapid, disappearance of these dramatic birds signaled to ornithologists that something was seriously wrong with the environment, but what? It took a great deal of clever biological detective work (see Ratcliffe, 1980) to link the decline of peregrine populations to the accumulation of these persistent pesticides up the food chain, resulting in eggshell thinning, reproductive failure, and (in extreme cases) direct poisoning of adult birds. Although some populations have now recovered, signaling a recovery in environmental quality, the species is still missing from many parts of its former range—some coastal populations in England, for instance. Nobody knows why.

2. Migratory Songbird Declines in North America

In North America, considerable concern is currently being expressed over widespread declines in summer migrant birds, particularly warblers. Unlike the so-called amphibian decline, nobody questions the phenomenon; just like the amphibian decline, nobody re-

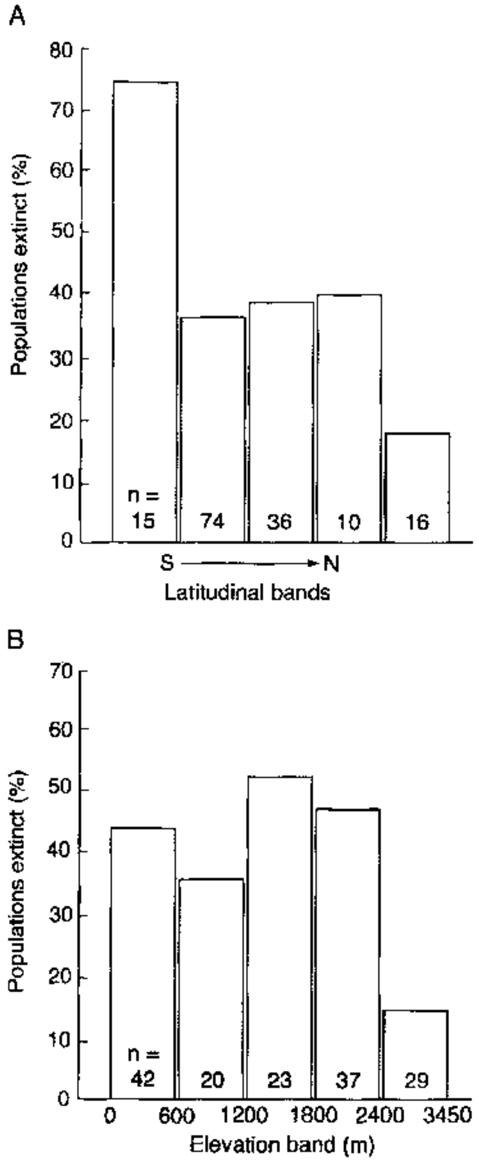


FIGURE 4 The fate of 151 previously recorded populations of Edith's checkerspot butterfly, *Euphydryas editha*, in western North America. The populations ranged from northern Mexico to southern Canada and were visited by Camille Parmisan and other biologists between 1992 and 1996. Populations that had disappeared because of habitat degradation (e.g., loss of usable host plants) were omitted from the analysis. Dividing the populations into five, evenly spaced latitudinal bands between 30° N and 53° N (A) reveals that significantly more southern populations have gone extinct than northern populations; sites where previously recorded populations still exist were, on average, 2° further north than sites where populations were extinct. Extinctions were also higher at lower altitudes (B) (*n* is the number of populations in each latitudinal or altitudinal band). Both results are consistent with the effects of global climate warming on the butterfly, leading to a northward and upward shift in its geographical range. Reproduced, with the permission of McMillan Journals Ltd, from *Nature* 282 (1996), page 766.

ally knows why it is happening. There is no doubting the data; many species are indeed declining very quickly, in as clear an indication as one wants that something is wrong with the environment, but what? Several possibilities exist, and they are unlikely to be mutually exclusive. One explanation focuses on the destruction of tropical forests in the birds' wintering areas. Another suggestion is that there are other unknown problems there or on the migration routes. A third possibility is extensive habitat fragmentation and urbanization in the breeding forests of the eastern seaboard. This human modification of the northeast forests markedly increases nest losses of migrant songbirds to jays, crows, cowbirds, and racoons, all species that thrive in the slipstream of urban humans.

3. Declines in Formerly Common Farmland Birds in Northwest Europe

In the intensively agricultural areas of northwest Europe—over the whole of lowland England, for example—a whole raft of formerly "common farmland birds" are also in steep decline (Tucker and Heath, 1994). They include skylarks (*Alauda arvensis*), European tree sparrows (*Passer montanus*), corn buntings (*Milaria calaudra*), gray partridges (*Perdix perdix*), and song thrushes (*Turdus philomelos*). Here the problem is now reasonably well understood, though many details remain unresolved. Modern farming is so efficient and clean that there is little for the birds to eat. Weeds are killed with herbicides, which remove both seeds and rich sources of insects that feed on the weeds. The crop itself is sprayed to remove insects and is harvested so efficiently that few seeds are spilled on the way. Modern farms are biodiversity deserts, an indication of the power of people to squeeze nature to the margins while apparently maintaining a green and pleasant land. If present trends continue, skylarks will be rare birds in Britain in 20 years.

III. SPECIES AS INDICATORS OF BIODIVERSITY

A. The Nature of the Problem

Common sense suggests that the known losses of plants and birds from European farmland will go hand-in-hand with much more poorly documented declines in many other, less familiar and cryptic taxa, from land snails to glowworms, and hoverflies to harvest spiders. In other words, changes in the distribution and abundance of well-known groups should serve as broad indicators of the status of, and changes in, a much wider sample of a region's flora and fauna. The assumption here is that birds (or other conspicuous species) might serve as biodiversity indicators—that is, as surrogates of overall biodiversity. But although it seems intuitively reasonable to use familiar, well-studied, and easily censused groups as indicators of what is happening to many other taxa, despite a great deal of research, the idea is actually contentious.

Following (but slightly modifying) the work of McGeoch (1998), we can define a biodiversity indicator as a group of taxa (e.g., genus, tribe, family, or order, or a selected group of species from a range of higher taxa) whose diversity (e.g., overall species richness, number of rare species, levels of endemism) reflects that of other higher taxa in a habitat, group of habitats, or geographic region. The idea is simple enough, and if it can be shown to work, it is important because biologists then have a relatively simple means of assessing overall biodiversity for purely scientific reasons, for setting conservation priorities, or for monitoring the effectiveness of conservation management.

B. Taxa That Have Been Suggested as Indicators of Biodiversity

The groups of organisms whose richness has been evaluated most thoroughly in the greatest number of places on earth are also the most familiar. The natural history sections of bookshops are dominated (sometimes exclusively) by volumes on plants and birds. If insects figure at all, butterflies will be on the top of the list, although there are fascinating differences between nations. Japan loves dragonflies. Birds, higher plants, butterflies, and dragonflies are all groups that occur in most places in the world but whose individual species are seldom so widespread. In much of the world they are also groups whose species are, relatively speaking, taxonomically well known and stable, readily identifiable, and have biologies that are well understood. They are easy to find, inventory, and count, and they are reasonably, but not overwhelmingly, diverse in any one place. These are all desirable attributes of groups that mightbe used as indicators of the diversity of many other, much less well known taxa—that is, as indicators of the overall biodiversity of a region.

Other groups have many of these same attributes but have not gained the same popularity, perhaps because often they are not also large bodied or perceived as being quite so attractive. The list of those that have been advocated as useful biodiversity indicators at one time or another is very long. It includes soil nematodes, moths, beetles galore (tiger, carabid, dung, and buprestid, to name but four), termites, fish, frogs, and snakes. Whatever the group, they must also have one further attribute—namely, that they genuinely indicate levels of biodiversity or at least some of the components of primary interest. The fact that the scientific literature contains suggestions for so many different possible indicators shows that there is little consensus on the matter. Many have been called, but few are chosen. Why? There are two, related, reasons. First, scientific knowledge on the degree of coincidence in patterns of biodiversity between different taxa is surprisingly poor. Second, as knowledge improves, coincidence between many taxa turns out to be much worse than people had imagined, or indeed hoped, would be the case.

C. Knowledge Is Poor Because of the Effort Required

Gathering information on the diversity of different groups of organisms, even in one place, is enormously time-consuming. Two examples illustrate the problem. To map the presence and absence of breeding birds (conspicuous, "easy" to find and to identify) in every 10 × 10 km grid-square in Britain and Ireland (there are 3672 squares) took more than half a million individual record cards, filled in by an army of amateur birdwatchers coordinated by professional ornithologists in the British Trust for Ornithology (BTO). The task took 4 years and about 100,000 hours of fieldwork (Gibbons *et al.*, 1993). Now imagine the effort required to do the same thing for all the other hundreds of different groups of organisms found in this one small corner of Europe. It has been done for a sample of taxa (we will return to what these data show in a moment), but many groups remain unmapped.

At a much smaller spatial scale in a tropical forest in Cameroon, a group of biologists attempted to measure the impacts of forest disturbance on just eight groups (birds, butterflies, flying beetles, canopy beetles, canopy ants, leaf-litter ants, termites, and soil nematodes). The birds and the butterflies took 50 and 150 scientist-hours, respectively, to survey. But the effort required climbed rapidly for smaller-bodied, more cryptic, less well-known groups—1600 hours for the beetles, 2000 for the termites, and 6000 for the nematodes (Lawton *et al.*, 1998). Despite the fact that this work as a whole took about five scientist-years, inventories for most groups that were surveyed were still only partial, and most taxa remained unexamined (fungi, higher plants, spiders, soil mites, collembola, earthworms, lizards, frogs, and mammals, to name some of the most conspicuous gaps).

Given this background, it is hardly surprising that biologists do not have a complete inventory of all the species that occur even in a single, moderately sized area (a field, small wood, or lake)—a so-called ATBI (All Taxa Biodiversity Inventory) (Oliver and Beattie, 1996). A moments thought will also show that to use one or two groups (for the sake of the argument, say birds and butterflies) as indicators of the richness of other taxa in fact requires *several* such areas to be investigated to properly test the hypothesis that high bird diversity (or any other single group) reflects a high diversity of many other groups.

Although progress has been made in this area over the past decade, considerable work remains to be done. Even in otherwise well-studied situations, many groups remain to be examined. Hence, at the present time, and effectively by default, some groups are being used as indicators of biodiversity, even though we cannot show categorically that the richness of one or more groups of organisms truly reflects the overall, or even a major portion of the overall, biodiversity of an area. As a result there is little consensus about what a "good" indicator group, or groups, might be, because there are too few hard data, from a range of habitats and geographic regions round the world, on which to draw firm conclusions. But as data slowly emerge, they are not encouraging for those who wish to use simple, single-taxon indicators of biodiversity.

D. Indicator Reliability

Where knowledge exists, it suggests that single or small numbers of taxa will usually be poor indicators of the biodiversity of other groups.

1. Tropical versus Temperate and Other Major Diversity Gradients

It would be wrong to think that there is no coincidence between patterns of diversity in different groups of organisms. Of course there is. In the broadest terms, it is axiomatic that most major terrestrial and freshwater groups are more species rich in the tropics than in temperate regions, at low elevations than at high ones, in forests than in deserts, and on large land masses than tiny islands. Whether you are a botanist, a bird-watcher, or a bug hunter, to find the most species it is generally advisable to head to hot and humid mainland tropics with lots of trees. It is easy to assume that there must therefore be reasonably good correlations between major diversity gradients for different groups. There can be, but even at this scale often there are not. Penguin diversity peaks in Antarctica, not the tropics, and there are many other examples of similar "reverse diversity gradients" that buck the average trend. On the eastern side of North America, the diversity of breeding warblers *increases* from south to north—suggesting that this conspicuous taxon, which is easy to identify (at least the breeding males!) and to survey, is probably highly unsuitable as an indicator of patterns of biodiversity in most other taxa (in which diversity typically *decreases* from south to north).

In the case of breeding North American warblers, we can spot the problem because we have enough information about the organisms involved. But the whole point about indicator taxa for biodiversity is that typically we will not be armed with, and indeed should not need, information about "other" groups; knowledge of the indicator taxon should suffice and be reliable. The evidence suggests otherwise.

2. Hot Spots

Major gradients in diversity aside, at similarly large scales an indicator group might be used to identify local geographic hot spots in the species richness of one or more other groups (peaks in the landscape of species richness) or to determine relative levels of richness in those other groups (hot spots versus all spots) (Gaston 1996b; Reid 1998). At the continental scale, the procedure has frequently been found to fail on both counts (Gaston 1996a), with mismatches between the occurrence of peaks in the richness of different groups being commonplace.

Across the United States and southern Canada, hot spots (local areas with unusually high diversity) overlap partially between some pairs of taxa (trees, tiger beetles, amphibians, reptiles, birds, and mammals), but the pattern is not a general one. Numbers of species in different large grid cells for two groups are often significantly positively correlated, for example, birds and tiger beetles or mammals and swallowtail butterflies. But these correlations are frequently weak, of rather limited predictive value, and in some cases explained by latitudinal gradients in diversity. In other words, although such correlations may sometimes enable a very general impression of the patterns in richness of one group to be obtained from the patterns in richness of another, their predictive powers are low.

These conclusions seem to hold at finer resolutions over more constrained areas. Thus, species-rich areas for different taxa in Britain (birds with butterflies, dragonflies, etc.) frequently do not coincide at a scale of 10 × 10 km squares (Pendergast *et al.*, 1993) (Fig. 5).

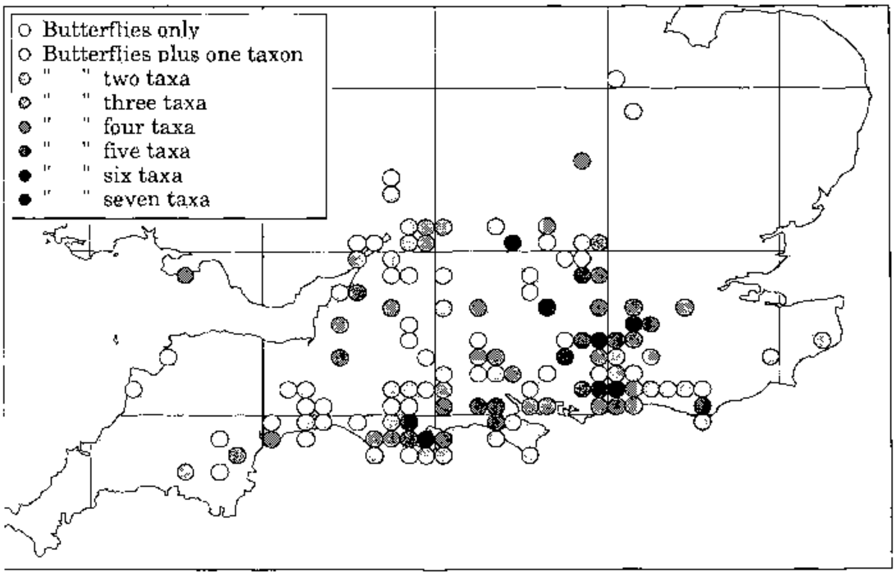


FIGURE 5 Coincidence between hot spots for butterflies and up to seven other taxa in Britain. Hot spots are unusually species-rich sites (here defined as the top 5 percentile in Britain). All of the most species-rich localities in Britain for butterflies lie in southern England. Increasingly dark shading indicates that butterfly hot spots coincide with hot spots for an increasing number of other taxa. Note that many butterfly hot spots are not unusually rich in any other species (open circles) and that only one locality (in southeast England, just in from the coast) is a hot spot for all eight taxa in this particular survey. The other taxa are breeding birds, dragonflies, moths, mollusks, aquatic higher plants, and liverworts (simple plants). Information from Prendergast *et al.* (1993), with additional data and figure kindly provided by John Prendergast.

Hot spots in this study are not distributed randomly, overlapping more often than expected by chance, but still at a low level. Likewise, different taxa are species poor or species rich in different areas of the Transvaal region of South Africa. At even finer scales, within the Cameroon forest mentioned earlier, disturbance impacted on the diversity of eight taxa in very different ways. All declined drastically in completely cleared areas, but intermediate levels of forest disturbance had very different effects on the diversity of different groups. As a result, changes in the diversity of one taxon could not be used to predict changes in the diversity of any other (Lawton *et al.*, 1998). A summary of these and related studies showing similar results is provided by Gaston (1996a, 1996b) and by Pimm and Lawton (1998).

3. A Commonsense Explanation

This lack of, or relatively feeble correlation between, species rich-areas for different groups of organisms makes the search for simple, robust, single-taxon indicators of overall biodiversity look increasingly like a lost cause. With hindsight, perhaps this emerging result is obvious (Reid, 1998). Major geographic gradients in biodiversity aside, within particular geographic regions or at smaller habitat scales, the conditions favoring one group of species may be hostile to another. Mollusks like it cool and wet, butterflies like it warm and sunny, and high bird diversity is more likely in tall vegetation than short vegetation, irrespective of weather. Commonsense natural history suggests that there is unlikely to be a single indicator taxon able to predict the diversity of all, or even a majority of others.

4. Rare Species and Endemic Species

Biologists and conservationists are often interested not only in patterns of species richness but also in the distribution of unusually rare species, or of endemic species. Do sites with unusual numbers of rare species frequently coincide across different taxa? Again, the answer seems to be no, or only weakly (Pimm and Lawton, 1998; Prendergast *et al.*, 1993), for the reasons just outlined.

Endemic species may be different (Bibby *et al.*, 1992). There is some evidence that areas rich in endemic birds (e.g., some tropical mountaintops or isolated islands) may also contain unusually large numbers of endemic species in other groups. However, rigorous data and analyses are few, and exceptions are easy to find. Lake Baikal has no endemic birds but supports an exceptionally rich, endemic invertebrate fauna and a unique, endemic freshwater seal.

E. Selecting Areas to Conserve Biodiversity: Conservation Planning and Reserve Selection Algorithms

1. Biodiversity Indicators and Conservation

The general lack of reliable indicator groups for biodiversity is undoubtedly unfortunate for scientists wishing to understand how life is distributed across the earth; the road to an atlas of biodiversity seems set to be a long one. In practice, it may actually prove somewhat less of a worry for one of the primary motivations in the search for indicators for biodiversity—namely, conservation planning. Networks of national parks and reserves are central planks in conservation, albeit alone they are insufficient to protect all species. Their establishment is one of the obligations placed on Parties to the Convention on Biological Diversity. A primary argument for using indicators of biodiversity is to determine the effectiveness of these protected area networks in capturing biodiversity, and the best ways in which they might be extended, in the face of stiff competition with alternative forms of land use.

Although hot spots of species richness coincide weakly for different taxa (see the previous section), if we turn the problem around and look at it from a different angle, an interesting picture emerges. Imagine that conservation priorities in Britain have been set by concentrating just on areas rich in birds (the bird hot spots). (Although the general view of many nations is that the British are mad about birds, the country's protected area network is not based solely on birds. We use the example simply to illustrate a point.) What we discover is that the hot spots for this one group tend to embrace a high proportion of the total species in other groups. Thus, the hot spots for breeding birds contain 87% of the breeding bird species in Britain, 100% of the butterflies, 92% of the dragonflies, 92% of the liverworts, and 94% of the aquatic plants. A reserve network established around hot spots for one group does a rather good job of ensuring that most species in other groups find a place in the extended ark of protected sites.

2. Reserve Selection Algorithms

Although this message is encouraging, it turns out that designing conservation networks simply on the basis of levels of species richness is extremely inefficient, at least if the goal is to capture representative samples of all taxa. The same species may occur repeatedly in different richness hot spots for a group. Conservationists do not have unlimited resources, and this duplication wastes money on purchasing, or managing, unnecessary land. On the other hand, some species, particularly the rare ones of primary conservation interest, may not occur in any richness hot spots at all. What is required is to identify those areas that constitute the greatest complementary species richness; the complementary part of an area's biota consists of those species unrepresented in another biota with which it is being compared. To do this, mathematically minded conservation biologists have developed powerful reserve selection algorithms that help to select sites with maximum efficiency, according to some predetermined criteria (Pressey *et al.*, 1993). The criteria may be to maximize the number of species, rare species, or endemic species in a proposed reserve network, at minimum cost, on a minimum area, closest to existing reserves, or what have you.

Echoing the conclusions of the previous section, the question then arises as to whether the patterns of complementarity of one group of organisms are congruent with those of another. The question is a new one, with few studies available to answer it. Across 50 forests of Uganda, which boasts more species for its size than almost any other country in Africa, and consistent with our earlier conclusions, there was little spatial congruence in the species richness of woody plants, large moths, butterflies, birds, and small mammals once differences in sampling effort were accounted for. However, sets of forests selected using complementarity determined for single taxa were generally similar to those for all other taxa and hence served to capture well the species richness in all these other groups (Howard *et al.*, 1998).

If these results generalize to other parts of the world, they send an encouraging message to conservation managers struggling to identify the best areas to set aside as reserves and parks. It says that a complementary and therefore efficiently selected chain of reserves based on a single indicator taxon (or perhaps two or three indicator taxa) may efficiently capture complementary sets of many other groups as well. Unfortunately, the Ugandan results are not supported by similar studies in the Transvaal, elsewhere in Africa (van Jaarsveld *et al.*, 1998). It may therefore be too soon to assume that we can find simple indicators for complementary reserve sets embracing many taxa as a means of conserving biodiversity.

IV. CONCLUSIONS

The term "indicator species" has three distinct meanings. They are a species, or group of species, that reflect the biotic or abiotic state of an environment; reveal evidence for, or the impacts of, environmental change; or indicate the diversity of other species, taxa, or entire communities within an area. The uses of indicator species in the first two senses of the word are very similar, differing largely in the fact that to indicate change, organisms need to be sampled more than once in the same place and in the same way. Using organisms to indicate the state of, and changes in, the environment has numerous tried and tested applications, from detecting pollution to monitoring recovery of formerly degraded habitats, at many scales, from local to global. The use of indicator species to predict the diversity of other, unstudied taxa for scientific or conservation reasons is much more contentious and may prove to be impossible with any degree of rigor.

See Also the Following Articles

ACID RAIN AND DEPOSITIONS • BIRDS, BIODIVERSITY OF • ECOTOXICOLOGY • ENDEMISM • ENVIRONMENTAL IMPACT, CONCEPT AND MEASUREMENT OF • GREENHOUSE EFFECT • HOTSPOTS • KEYSTONE SPECIES • PALEOECOLOGY

Bibliography

Bibby, C. J., Collar, N. J., Crosby, M. J., Heath, M. F., Imboden, C., Johnson, T. H., Long, A. J., Stattersfield, A. J., and Thirgood, S. J. (1992). Putting Biodiversity on the Map: Priority Areas for Global Conservation. International Council for Bird Preservation, Cambridge.

Blaustein, A. R., and Wake, D. B. (1995). The puzzle of declining amphibian populations. *Sci. Am.* April, 56–61.

Erisman, J. W., and Draaijers, G. P. J. (1995). Atmospheric Deposition in Relation to Acidification and Eutrophication. Elsevier, Amsterdam.

Gaston, K.J. (1996a). Biodiversity—Congruence. *Prog. Phys. Geog.* **20**, 105–112.

Gaston, K. J. (1996b). Spatial covariance in the species richness of higher taxa. In *Aspects of the Genesis and Maintenance of Biological Diversity* (M. E. Hochberg, J. Clobert, and R. Barbault, Eds.), pp. 221–242. Oxford University Press, Oxford.

Gibbons, D. W., Reid, J. B., and Chapman, R. A. (Eds.) (1993). *The New Atlas of Breeding Birds in Britain and Ireland: 1988–1991*. Poyser, London.

Howard, P. C., Viskanic, P., Davenport, T. R. B., Kigenyi, F. W., Baltzer, M., Dickinson, C. J., Lwanga, J. S., Matthews, R. A., and Balmford, A. (1998). Complementarity and the use of indicator groups for reserve selection in Uganda. *Nature* **394**, 472–475.

Huntley, B., and Birks, H. J. B. (1983). *An Atlas of Past and Present Pollen Maps for Europe: 0–13000 Years Ago*. Cambridge University Press, Cambridge.

van Jaarsveld, A. S., Fretag, S., Chown, S. L., Muller, C., Koch, S., Hull, H., Bellamy, C., Krüger, M., Endródy-Younga, S., Mansell, M. W., and Scholtz, C. H. (1998). Biodiversity assessment and conservation strategies. *Science* **279**, 2106–2108.

Lawton, J. H., Bignell, D. E., Bolton, B., Bloemers, G. F., Eggleton, P., Hammond, P. M., Hodda, M., Holt, R. D., Larsen, T. B., Mawdsley, N. A., Stork, N. E., Srivastava, D. S., and Watt, A. D. (1998). Biodiversity inventories, indicator taxa and the effects of habitat modification in tropical forest. *Nature* 391, 72–76.

McGeoch, M. A. (1998). The selection, testing and application of terrestrial insects as bioindicators. *Biol. Rev.* 73, 181–201.

Oliver, I., and Beattie, A. J. (1996). Designing a cost-effective invertebrate survey: A test of methods for rapid assessment of biodiversity. *Ecol. Appl.* 6, 594–607.

Pimm, S. L., and Lawton, J. H. (1998). Planning for biodiversity. *Science* 279, 2068–2069.

Prendergast, J. R., Quinn, R. M., Lawton, J. H., Eversham, B. C., and Gibbons, D. W. (1993). Rare species, the coincidence of diversity hotspots and conservation strategies. *Nature* **365**, 335–337.

Pressey, R. L., Humphries, C. J., Margules, C. R., Vane-Wright, R. I., and Williams, P. H. (1993). Beyond opportunism: Key principles for systematic reserve selection. *TREE* **8**, 124–128.

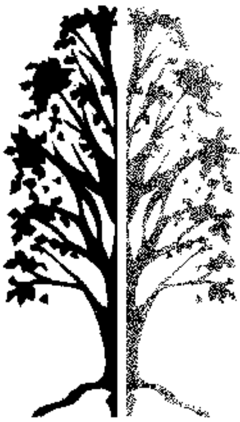
Ratcliffe, D. A. (1980). *The Peregrine Falcon*. Poyser, Calton.

Reid, W. V. (1998). Biodiversity hotspots. *TREE* **13**, 275–280.

Rosenberg, D. M., and Resh, V. H. (Eds.) (1993). *Freshwater Biomonitoring and Benthic Macroinvertebrates*. Chapman & Hall, New York.

Shaw, I. C., and Chadwick, J. (1998). *Principles of Environmental Toxicology*. Taylor & Francis, London.

Tucker, G. M., and Heath, M. F. (1994). *Birds in Europe: Their Conservation Status*. BirdLife International, Cambridge.



INDIGENOUS PEOPLES, BIODIVERSITY AND

Victor M. Toledo
Institute of Ecology, National University of Mexico (UNAM)

GLOSSARY

bio-cultural axiom Recognition that biological and cultural diversity are mutually dependent and geographically coterminous.

cultural diversity Variety of human groups distinguished through beliefs, lifeways, dress, food, languages, sexual behavior, forms of productive organization, art, and conceptions of nature.

endemic languages Languages that are restricted to a single country and, like their species counterparts, hold a high percentage of the unique traits in human language.

ethnoecology Interdisciplinary study that explores how nature is perceived by human groups through a screen of beliefs and knowledge and how humans, in terms of their images and symbols, use and/or manage natural resources.

indigenous peoples Those who are the "original" or oldest inhabitants of an area or region, or who have lived in a traditional homeland for many generations, usually many centuries.

INDIGENOUS PEOPLES NUMBER BETWEEN 300 AND 700 MILLION. They are inhabitants of practically every major biome of the earth and especially of the least disturbed terrestrial and aquatic ecosystems. This article is based on an exhaustive review of recently published data, and stresses the strategic importance of indigenous peoples in the maintenance and conservation of the world's biodiversity. Four main links between biodiversity and indigenous peoples are examined: the correlation between biological richness and cultural diversity on both geopolitical and biogeographic terms; the strategic importance of indigenous peoples in the biomass appropriation; the remarkable overlap between indigenous territories and the world's remaining areas of high biodiversity; and the importance of indigenous views, knowledge, and practices in biodiversity conservation. The article concludes by emphasizing the urgent need for recognizing a new biocultural axiom: that global biodiversity can only be effectively preserved by preserving the diversity of human cultures, and vice versa.

Encyclopedia of Biodiversity, Volume 3

I. INTRODUCTION

Biodiversity as a word and concept originated in the field of conservation biology. However, as Alcorn (1994: 11) states, "while proof of conservation success is ultimately biological, conservation itself is a social and political process, not a biological process. An assessment of conservation requires therefore an assessment of social and political institutions that contribute to, or threaten, conservation." One of the main social aspects related to biodiversity is, undoubtedly, the world's indigenous peoples.

Scientific evidence shows that virtually every part of the planet has been inhabited, modified, and manipulated throughout human history. Although they appear to be untouched, many of the last tracts of wilderness are inhabited and have been so for millennia. Indigenous peoples live in and have special claims to territories that, in many cases, harbor exceptionally high levels of biodiversity. On a global basis, human cultural diversity is associated with the remaining concentrations of biodiversity. Both cultural diversity and biological diversity are endangered.

Given the foregoing, this article offers a review of the multiple importance of indigenous peoples and makes the point that valuable, local-specific views, knowledge, and practices are used by indigenous peoples who have relied for centuries on the maintenance of biodiversity. Indigenous peoples are often classified as impoverished or treated as invisible. However, in the final analysis, they hold the key to successful biodiversity conservation in most of the biologically richest areas of the world.

II. INDIGENOUS PEOPLES

TABLE I

Estimated Population of the World's Indigenous Peoples[a]

Region	Number of cultural groups	Population
North America	250	3,500,000
Latin America and the Caribbean	800	43,000,000
Former Soviet Union	135	40,000,000
China and Japan	100	67,000,000
The Pacific	1,273	2,000,000
Southeast Asia	900	30,000,000
South Asia	700	100,000,000
Australia and New Zealand	250	550,000
Africa	2010	50,000,000
Total	6418	336,050,000

[a] Sources: Burger (1987); Hitchcock (1994).

Indigenous peoples number between a minimum of 300 million (Table I) and a maximum of 700 million (Harmon, 1995; the reason for this wide range is discussed in the following). They live in about 75 of the world's 184 countries and are inhabitants of practically each major biome of the earth. Indigenous peoples—also called tribal, aboriginal, or autochthonous peoples, national minorities, or first peoples—are best defined by using several criteria. Indigenous peoples may have all or some of the following criteria: (a) are the descendants of the original inhabitants of a territory that has been overcome by conquest; (b) are "ecosystem peoples," such as shifting or permanent cultivators, herders, hunters and gatherers, fishers, and/or handicraft makers, who adopt a multiuse strategy of appropriation of nature; (c) practice a small-scale, labor-intensive form of rural production that produces little surplus and has low energy needs; (d) do not have centralized political institutions, organize their life at the level of the community, and make decisions on a consensus basis; (e) share a common language, religion, moral values, beliefs, clothing style, and other identifying characteristics, as well as a relationship to a particular territory; (f) have a different worldview, consisting of a custodial and nonmaterialist attitude to land and natural resources based on a symbolic interchange with the natural universe; (g) are subjugated by a dominant culture and society; and (h) consist of individuals who subjectively consider themselves to be indigenous.

It is possible to find indigenous peoples carrying out many different activities of use and management of the planet's ecosystems: as forest-dwellers in the tropical lowlands or in the mountains, as pastoralists in savannas and other grasslands, or as nomadic or seminomadic hunters and gatherers in forests, prairies, and deserts. In addition, fishing is the principal economic activity and source of food for several million coastal and island dwellers, as well as for many indigenous peoples inhabiting margins of rivers.

Large numbers of indigenous peoples are, however, peasant producers and therefore can be indistinguishable from the nonindigenous peoples living nearby. In the Andean and Mesoamerican countries of Latin America, for instance, indigenous peoples farm like mestizo peasants. Similarly, in India distinctions between scheduled (government recognized) tribes and nontribal peoples cannot be made solely on the basis of productive activities. In these and other many cases, nonindigenous peasants and indigenous peoples produce the same crops with the same farming methods. Because in numerous countries many mestizo peasants are direct descendants of the indigenous peoples and retain most of their cultural traits, it has been pointed out that a broader definition of indigenous peoples might increase the real numbers. Thus, by considering characteristics other than language, it is possible to enlarge the number of people classified as indigenous in the contemporary world. Some authors, such as J. Burger (1987), think that the number of indigenous people may actually be double that previously estimated. Thus, in the contemporary world there may be as many as 600 million indigenous peoples. Burger's estimation echoes the figure of over 700 million given by Harmon (1995) as the total number of people speaking some of the world's 5635 "endemic languages," which are languages restricted to only one nation. Endemicity in language can be linked to small-scale societies and, therefore, to indigenous peoples.

Based on the percentage of the total population identified as being indigenous, it is possible to recognize a group of selected nations with a strong presence of these peoples: Papua New Guinea (77%), Bolivia (70%), Guatemala (47%), Peru (40%), Ecuador (38%), Myanmar (33%), Laos (30%), Mexico (12%), and New Zealand (12%). On the other hand, the absolute number of people recognized as indigenous allows one to identify nations with high indigenous populations, such as India and China.

III. BIOLOGICAL DIVERSITY AND DIVERSITY OF CULTURES

On a global basis, human cultural diversity is associated with the remaining concentrations of biodiversity. In fact, evidence exists of remarkable overlaps between global mappings of the world's areas of high biological richness and areas of high diversity of languages, which is the single best indicator of a distinct culture. According to Harmon (1996), "Species and languages are not just comparable on an abstract, conceptual level. There is also a striking pattern of congruity in the geographical distribution of the two. For instance, many countries with high numbers of endemic species also have many endemic languages."

Measured by spoken language, all the world's people belong to between 5000 and 7000 cultures. It is estimated that 4000 to 5000 of these are indigenous cultures. Thus, indigenous peoples account for as much as 80 to 90% of the world's cultural diversity. On the basis of the inventories done by linguists, we can draw up a list of the regions and countries with the greatest degree of cultural diversity in the world. According to *Ethnologue*, the best existing catalog of the world's languages, there is a total of 6703 languages (mostly oral), 32% of which are found in Asia, 30% in Africa, 19% in the Pacific, 15% in the Americas, and 3% in Europe (Grimes, 1996). Only twelve countries account for 54% of all human languages. These countries are Papua New Guinea, Indonesia, Nigeria, India, Australia, Mexico, Cameroon, Brazil, Zaire, Philippines, United States, and Vanuatu (Table II).

On the other hand, according to the most recent and detailed analysis of biodiversity on a country-by-

TABLE II

Top 25 Countries by Number of Endemic Languages[a]

1. *Papua New Guinea (847)
2. *Indonesia (655)
3. Nigeria (376)
4. *India (309)
5. *Australia (261)
6. *Mexico (230)
7. Cameroon (201)
8. *Brazil (185)
9. *Zaire (158)
10. *Philippines (153)
11. *United States (143)
12. Vanuatu (105)
13. Tanzania (101)
14. Sudan (97)
15. *Malaysia (92)
16. Ethiopia (90)
17. *China (77)
18. *Peru (75)
19. Chad (74)
20. Russia (71)
21. Solomon Islands (69)
22. Nepal (68)
23. *Colombia (55)
24. Côte d'Ivoire (51)
25. Canada (47)

[a] After Harmon (1996). Asterisks indicate "megadiversity" countries according to Mittermeier and Goettsch-Mittermeier (1997).

country basis (Mittermeier and Goettsch-Mittermeier, 1997), there are, similarly, 12 countries that house the highest numbers of species and endemic species (Table III). This assessment was based on the comparative analysis of eight main biological groups: mammals, birds, reptiles, amphibians, freshwater fishes, butterflies, tiger beetles, and flowering plants. The nations considered to be "megadiversity" countries are: Brazil, Indonesia, Colombia, Australia, Mexico, Madagascar, Peru, China, Philippines, India, Ecuador, and Venezuela.

Thus, the relationship between cultural diversity and biological diversity stands out in global statistics: 9 of the 12 main centers of cultural diversity (in terms of number of languages) are also in the roster of biological megadiversity nations and, reciprocally, 9 of the countries with the highest species richness and endemism are also in the list of the 25 nations with the highest number of endemic languages (Harmon, 1996, see Table II and Table III).

The links between biological and cultural diversity can also be illustrated by using the data of Global 200, a program of the World Wide Fund for Nature (WWF) that was developed as a new strategy to identify conservation priorities based on an ecoregional approach. As part of this program, WWF identified a list of 233 terrestrial, freshwater, and marine biological ecoregions representative of Earth's richest diversity of species and habitats. A preliminary analysis conducted by the People and Conservation Unit of WWF regarding the presence of indigenous peoples in the 136 terrestrial ecoregions of Global 200 revealed interesting patterns. As shown in Table IV, nearly 80% of the terrestrial ecoregions are inhabited by one or more indigenous peoples, and half of the world's 3000 indigenous groups, as

TABLE III

Top 12 Countries by Number of Species (Richness) and Endemics (Endemism)[a]

	Biological diversity		
Country	Richness	Endemism	Both
Brazil	1	2	1
*Indonesia	3	1	2
*Colombia	2	5	3
*Australia	7	3	4
*Mexico	5	7	5
Madagascar	12	4	6
*Peru	4	9	7
*China	6	11	8
*Philippines	14	6	9
*India	9	8	10
Ecuador	8	14	11
Venezuela	10	15	12

[a] Calculated for the following biological groups: mammals, birds, reptiles, amphibians, freshwater fishes, butterflies, tiger beetles, and flowering plants. (Source: Mittermeier and Goettsch-Mittermeier (1997). Asterisks indicate countries included in the list of the 25 nations with the highest number of endemic languages (See Table II; Harmon, 1996).

TABLE IV

Indigenous Peoples (IP) in Global 200 Terrestrial Ecoregions Considered as Priority Areas by the World Wide Fund for Nature[a]

Region	Ecoregions	Ecoregions with IP	%	Total IP in the World[b]	Number of IP in ecoregions	%
World	136	108	79	3000	1445	48
Africa	32	25	78	983	414	42
Neotropics	31	25	81	470	230	51
Nearctic	10	9	90	147	127	86
Asia and Pacific (Indo-Malayan)	24	21	88	298	225	76
Oceania	3	3	100	23	3	13
Palearctic	21	13	62	374	111	30
Australasia	15	12	80	515	335	65

[a] Source: WWF International, People and Conservation Unit, unpublished report, August, 1998.

[b] These figures, utilized by WWF International, underestimate the number of indigenous peoples given by several analysts. For example, Harmon (1996) reported a total of 5,635 "endemic languages," which can be considered as socially equivalent to indigenous peoples.

estimated by WWF, are inhabitants of these ecoregions. All of the regions, except the Palearctic region, have indigenous peoples living in 78% or more of their identified ecoregions.

IV. BIODIVERSITY AND BIOMASS APPROPRIATION: THE ROLE OF INDIGENOUS PEOPLES

Biodiversity conservation cannot be separated from natural resources utilization. The human appropriation of natural materials includes minerals, water, shelter and fiber materials, solar energy, and, principally, living organisms (biomass) from ecosystems. World statistics indicate that almost half the humans on the planet are still engaged in the direct appropriation of natural resources. This appropriation is carried out by a myriad of rural or primary producers through the management of terrestrial, marine, and freshwater ecosystems.

Forty-five percent of the total human population has been recorded by the United Nations Food and Agriculture Organization (FAO) as agricultural population (FAO, 1991). It can be estimated that between 60 and 80% of this agricultural population is represented by small-scale, solar-energized productive units based on a multiuse management of nature (Toledo, 1990). In fact, the statistical record shows that by 1990 around 1.2 billion rural people were practicing agricultural activities on areas of 5 hectares or less. This figure coincides with the last available world census of agriculture by the FAO in 1970, when more than 80% of all reported holdings were smaller than 5 hectares (ha). A similar pattern is found in the world's fisheries, where more than 90% are small-scale, artisanal operators acting in a great variety of coastal habitats.

Most of these small-scale farmers and fishers develop their production activities not as socially isolated households but as familial nuclei belonging to specific village communities, many of which, in turn, correspond to cultures that can be considered as indigenous. Moreover, within the core of these community-based producers, those identified as indigenous people also carry out the biomass extraction at the lowest level in their local ecosystems. Called "ecosystem people" by some authors, such as R. F. Dasmann and M. Gadgil, these producers subsist by appropriating a diversity of biological resources from their immediate vicinity. Their quality of life is therefore intimately linked to the maintenance of certain levels of local biodiversity (Gadgil, 1993, and see the following). As a consequence, they are productive actors in little transformed habitats of the world, and include forest and sea dwellers, slash-and-burn agriculturalists, some 25–30 million nomadic herders or pastoralists (in East Africa, the Sahel, and the Arabian peninsula), most of the world's 15–21 million fishers, and all of the half a million hunters and gatherers still recognized as citizens of the contemporary world.

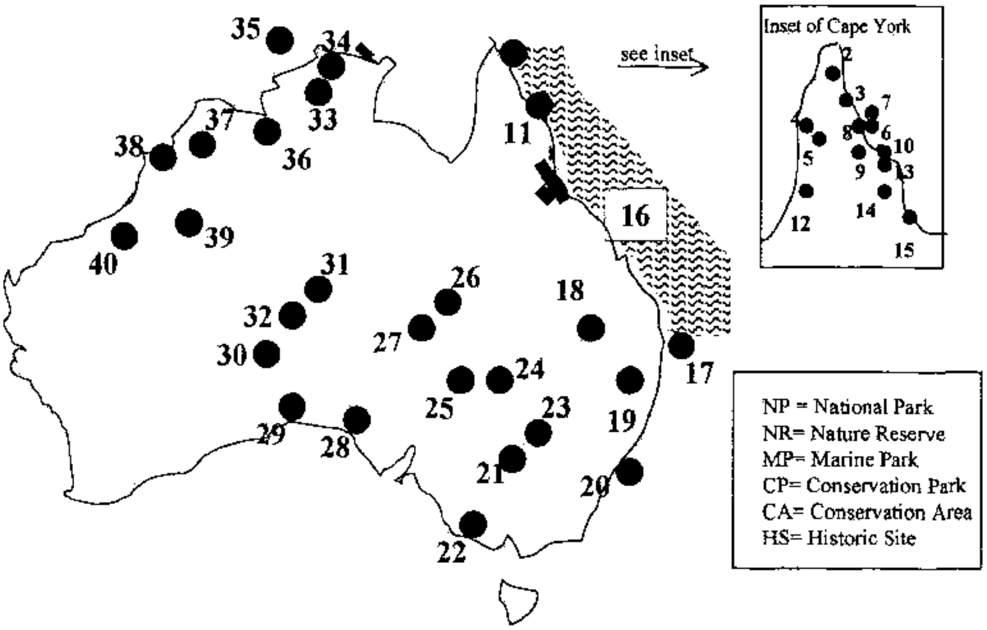
In conclusion, indigenous peoples represent the fraction of human appropriators of biomass that cause the lowest ecological impacts. They generally live in what may be termed "frontier lands" or "refuge regions," in other words, remote areas of great "wilderness" where the structure, not the components, of original ecosystems remains more or less untouched. In many cases, these lands and waters are untamed, unknown, unowned, and as yet unclaimed.

V. BIODIVERSITY AND INDIGENOUS PEOPLE'S LANDS AND WATERS

Indigenous peoples occupy a substantial share of the world's little disturbed tropical and boreal forests, mountains, grasslands, tundra, and desert, along with large stretches of coastline and nearshore waters (including mangroves and coral reefs) (Durning, 1993). The importance of indigenous territories to biodiversity conservation is therefore evident.

In fact, indigenous peoples control, legally or not, immense areas of natural resources. Among the most remarkable examples are the Inuit people (formerly known as Eskimo), who govern a region covering one-fifth of the territory of Canada (222 million ha), the indigenous communities of Papua New Guinea, whose lands represent 97% of the national territory, and the tribes of Australia with nearly 90 million ha (Fig. 1). Although numbering only above 250,000, the Amerindians of Brazil possess an area of over 100 million ha, mainly in the Amazon Basin, distributed in 565 territories (Fig. 2 and Table V). Nearly 60% of the priority areas in central and southern Mexico recommended for protection are also inhabited by indigenous peoples (Fig. 3), and half of the 30,000 rural communities are located in the 10 most biologically rich states of the Mexican territory. In summary, on a global scale it is estimated that the total area under indigenous control probably reaches between 12 and 20% of the earth's land surface (Stevens, 1997).

The best example of notable overlaps between indigenous peoples and biologically rich areas is the case of tropical humid forests. In fact, there is a clear correspondence between areas of remaining tropical forests and



1. Jardine River NP
2. Iron Range NP
3. Forbes Islands NP
4. Archer Bend NP
5. Rokeby-Croll NP
6. Flinders Island Group NP
7. Clack Islands
8. Cliff Islands
9. Lakefield NP
10. Starcke NP
11. Cape Melville NP
12. Mitchell and Alice Rivers NP
13. Mt. Webb NP
14. Cedar Bay
15. Mossman Gorge NP
16. Great Barrier Reef MP
17. Fraser Island NP
18. Carnarvon Gorge NP
19. Mt. Yarrowych NP
20. Jervis Bay NR
21. Lake Mungo NP
22. Coorong NP
23. Mt. Grenfell HS
24. Mootwingee NP
25. Gammon Ranges NP
26. Simpson Desert NP
27. Witjara NP
28. Yumbarra CA
29. Nullarbor NP
30. Unnamed CP
31. Watarraka NP
32. Uluru-Kata Tjuta NP
33. Nitmiluk NP
34. Kakadu NP
35. Gurig NP
36. Purnululu NP
37. Proposed Buccaneer Archipelago MP
38. Karlamilyi NP
39. Rudall River NP
40. Karijini NP

FIGURE 1 Terrestrial and marine protected areas with significant involvement or interest of indigenous peoples in Australia. (Adapted from data from the Australian federal government.)

the presence of indigenous peoples in Latin America, the Congo Basin in Africa, and several countries of tropical Asia, such as the Philippines, Indonesia, and Papua New Guinea. The strong presence of indigenous peoples in Brazil, Indonesia, and Zaire alone is remarkable, as these countries account for 60% of all the tropical forest in the world.

In Latin America, this geographical relationship has been strikingly verified for the Central American countries by a National Geographic Society map produced by a project headed by Mac Chapin in 1992. The same pattern can be found in the tropical humid areas of Mexico inhabited by 1.6 million indigenous people, and for many regions of the Amazonia Basin (see the case of Brazil in Fig. 2). It has been estimated that in Amazonia more than 1 million indigenous people of eight countries possess over 135 million ha of tropical forests (Davis and Wali, 1994).

Many temperate forests of the world also overlap with indigenous territories, for example, in India (Fig. 4), Myanmar, Nepal, Guatemala, the Andean countries (Ecuador, Peru, and Bolivia), and Canada. Furthermore, over 2 million islanders of the South Pacific, most of whom are indigenous peoples, continue fishing and harvesting marine resources in high-biodiversity areas (such as coral reefs).

VI. BIODIVERSITY AND ETHNOECOLOGY: INDIGENOUS VIEWS, KNOWLEDGE, AND PRACTICES

Biodiversity is a very broad concept that refers to the variety of landscapes, ecosystems, species, and genes, including the associated functional processes. Therefore, the maintenance and conservation of biodiversity demand efforts on these four levels. The first level is oriented to preservation of assemblies of "ecosystems," whereas the second level focuses on protection of habitats in which the populations of species live. At the species level, most biodiversity knowledge is of large

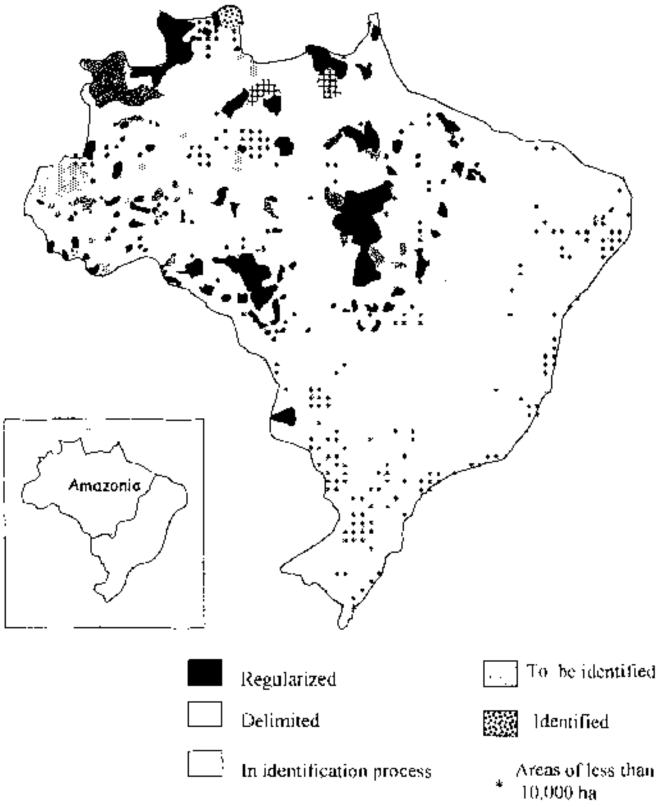


FIGURE 2 Geographical location of indigenous territories in Brazil, according to their legal situation and size area. Note the large tracts under indigenous control in the Amazonian region, the core of Brazilian biological richness. (Adapted from the map Terras Indígenas do Brasil, Instituto Socioambiental, São Paulo, Brazil.)

plants and animals such as flowering plants and vertebrates. Much of the extent of diversity of smaller plants and animals remains to be inventoried and protected. Although most biological diversity is constituted by wild plants and animals, an important subset involves the diversity among domesticated organisms. In this fourth level, interest focuses on the conservation of genetic variation in crops and domesticated animals.

TABLE V

Legal Situation of Indigenous Territory in Brazil (as of November, 1997)

Legal situation[a]	No. of indigenous areas	Area (ha)	%
Not identified	74	2,749,000	2.6
To be identified	96	4,983,578	4.92
Interdicted	5	8,897,066	8.88
Identified	12	1,998,117	1.97
Delimited	67	19,963,673	19.86
Demarcated and confirmed	73	14,816,728	14.77
Regularized	238	47,093,429	47.00
Total	565	100,501,591	100.00

[a] According to the National Indian Foundation (FUNAI) of Brazil.

This section examines the potential role of indigenous peoples in biodiversity conservation from an ethnoecological perspective. Ethnoecology can be defined as the interdisciplinary study of how nature is perceived by human groups through a screen of beliefs and knowledge, and how humans, through their symbols, use and/or manage natural resources. Thus, by focusing on the cosmos (the belief system or cosmovision), the corpus (the whole repertory of knowledge or cognitive systems), and the praxis (the set of practices), ethnoecology offers an integrative approach to the study of the process of human appropriation of nature (Toledo, 1992). This approach allows one to recognize the value of the belief–knowledge–practice complex of indigenous peoples in relation to the conservation of biodiversity.

A. The Cosmos

For indigenous peoples, land and in general nature have a sacred quality that is almost absent from Western thinking. Land is revered and respected and its inalienability is reflected in virtually every indigenous cosmovision. Indigenous people do not consider the land as merely an economic resource. Under indigenous cosmovisions, nature is the primary source of life that nourishes, supports, and teaches. Nature is, therefore, not only a productive source but also the center of the universe, the core of culture, and the origin of ethnic identity. At the heart of this deep bond is the perception that all living and nonliving things and natural and social worlds are intrinsically linked (the reciprocity principle). Of particular interest is the research done by several authors (G. Reichel-Dolmatoff, E. Boege, P. Descola, C. van der Hammen, and K. Arhem) on the role played by the cosmology of several indigenous groups as a mechanism regulating the use and management of natural resources. In the indigenous cosmovision, each act of appropriation of nature must be negotiated with all existing things (living and nonliving) through different mechanisms such as agrarian rituals and shamanic acts (symbolic exchange). Humans are thus seen as a particular form of life participating in a wider community of living beings regulated by a single and totalizing set of rules of conduct.

B. The Corpus

Indigenous societies house a repertory of ecological knowledge that generally is local, collective, diachronic,

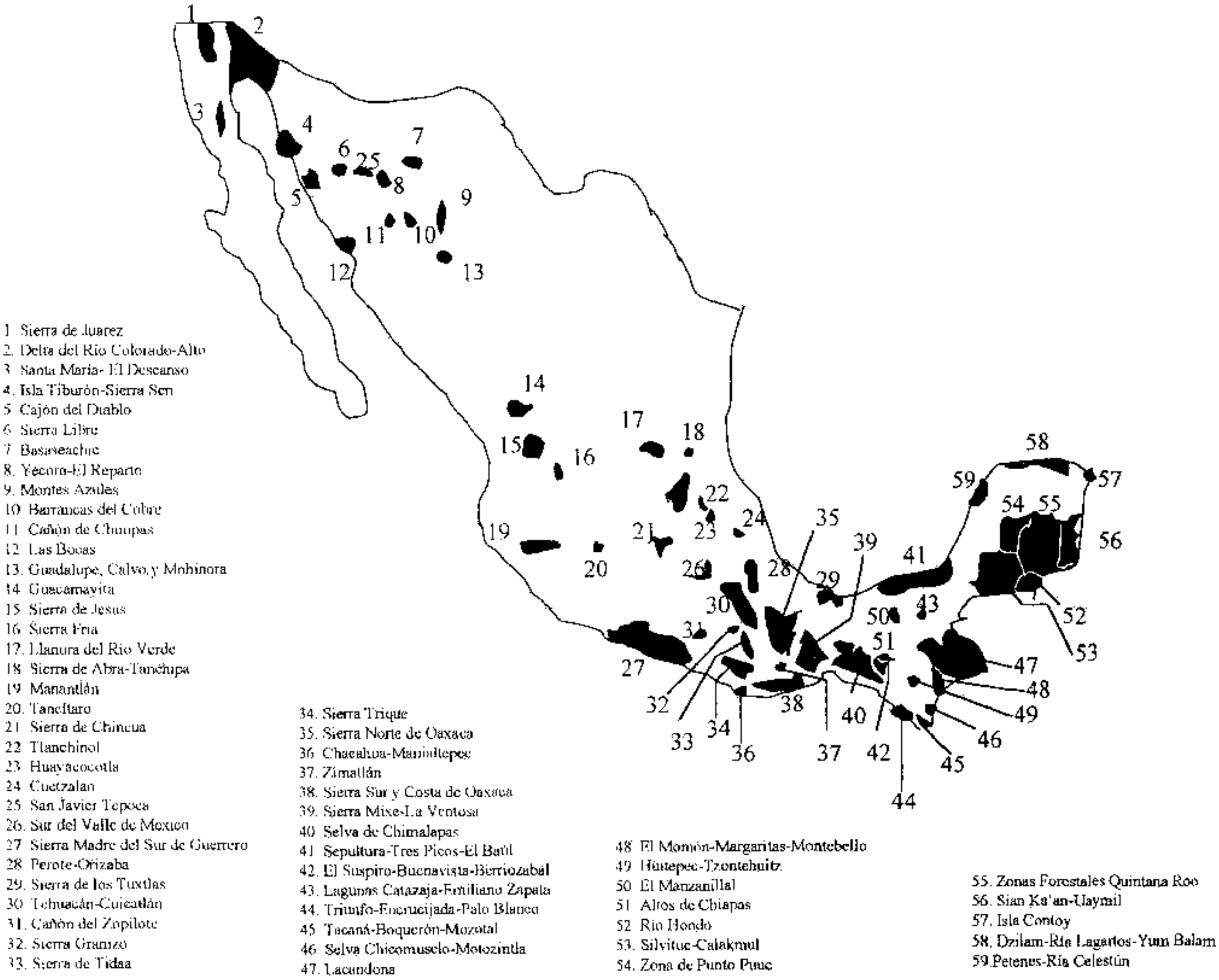


FIGURE 3 Geographical location of priority areas recommended by the Comisión Nacional para el Estudio y Uso de la Biodiversidad (CONABIO) of Mexico that overlap with territories of indigenous communities. Note the high number of overlapping areas in the central and southern portions of Mexico, where most of the biological richness of the country is concentrated. (Modified from CONABIO's map on priority areas for conservation, 1996.)

and holistic. In fact, because indigenous peoples possess a very long history of resource-use practice, they have generated cognitive systems on their own circumscribed natural resources, which are transmitted from generation to generation. The transmission of this knowledge is done through language, hence the corpus is generally an unwritten knowledge. Therefore, memory is the most important intellectual resource among indigenous cultures.

This body of knowledge is the expression of a certain personal wisdom and, at the same time, of a collective creation, that is to say, a historical and cultural synthesis turned into reality in the mind of an individual producer. For this reason, the corpus contained in a single producer's mind expresses a repertoire that is a synthesis of information from at least four sources: (a) the experience accumulated over historical time and transmitted from generation to generation by a certain cultural group; (b) the experiences socially shared by the members of a generation or cohort; (c) the experience shared in the household or the domestic group to which the individual belongs; and (d) the personal experience, particular to each individual, achieved through the repetition of the annual cycles (natural and productive), enriched by the perceived variations and unpredictable conditions associated with them.

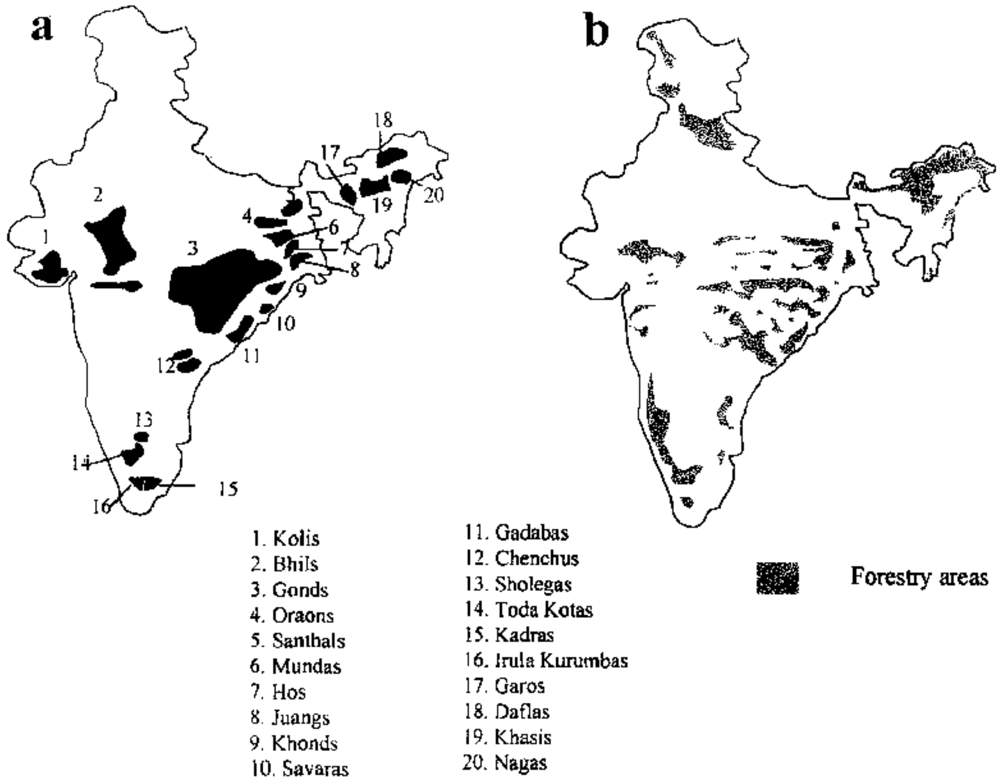


FIGURE 4 Geographical location of the main 20 indigenous groups (a) and principal forestry areas (b) of India. Although the long history of migrations of peoples makes it difficult to distinguish the indigenous peoples in India, there are about 100 million people considered by the government as "scheduled tribes" speaking over 300 languages. These groups are generally residents of remote hilly or forested areas. (Modified from The State of India's Environment 1984–85.)

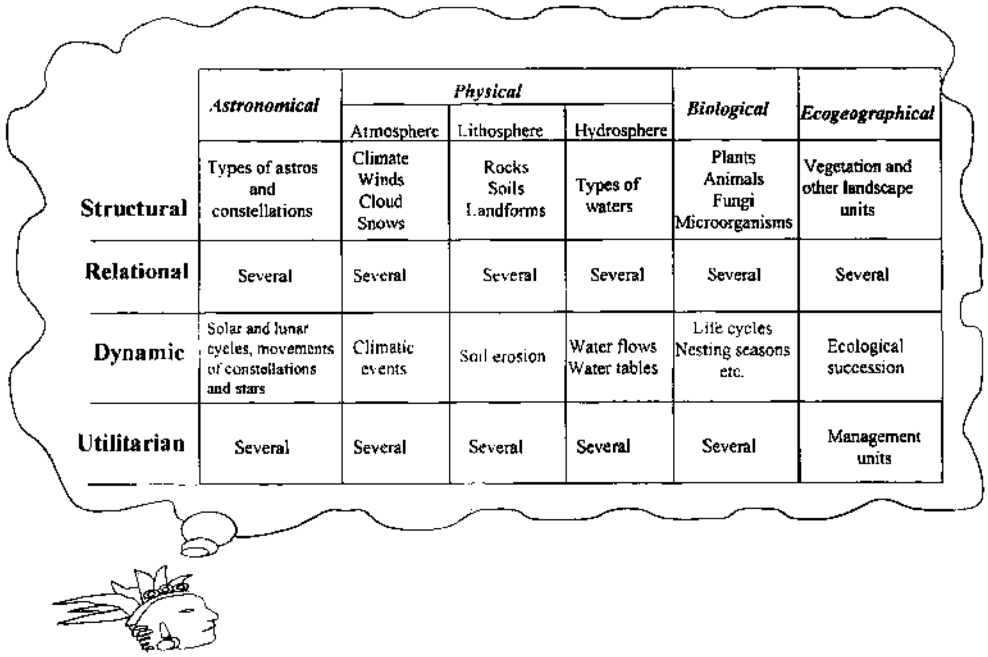
Thus, indigenous ecological knowledge is normally restricted to the immediate environments and is an intellectual construction resulting from a process of accumulation of experiences over both historical time and social space. These three main features of indigenous ecological knowledge—being local, diachronic, and collective—are complemented by a fourth characteristic, namely, the holistic.

Indigenous knowledge is holistic because it is intricately linked to the practical needs of use and management of local ecosystems. Although indigenous knowledge is based on observations on a rather restricted geographic scale, it must provide detailed information on the variety of scales represented by the concrete landscapes in which natural resources are used and managed. As a consequence, indigenous minds not only possess detailed information about species of plants, animals, fungi, and some microorganisms, but they also recognize many types of minerals, soils, waters, snows, landforms, vegetations, and landscapes.

Similarly, indigenous knowledge is not restricted to the structural aspects of nature, which are related to the recognition and classification (ethnotaxonomies) of elements or components of nature, but also encompasses dynamic (which refers to patterns and processes), relational (linked to relationships between or among natural elements or events), and utilitarian dimensions of natural resources. As a result, it is possible to integrate a cognitive matrix (Fig. 5) that certifies the holistic character of indigenous knowledge and serves as a methodological framework for ethnoecological research (Toledo, 1992).

C. The Praxis

Indigenous societies generally subsist by appropriating a diversity of biological resources from their immediate vicinity. Thus, subsistence of indigenous peoples is based more on ecological exchanges (with nature) than on economic exchanges (with markets). They are therefore forced to adopt survival mechanisms that guarantee an uninterrupted flow of goods, materials, and energy



	Astronomical	*Physical*			*Biological*	*Ecogeographical*
		Atmosphere	Lithosphere	Hydrosphere		
Structural	Types of astros and constellations	Climate Winds Cloud Snows	Rocks Soils Landforms	Types of waters	Plants Animals Fungi Microorganisms	Vegetation and other landscape units
Relational	Several	Several	Several	Several	Several	Several
Dynamic	Solar and lunar cycles, movements of constellations and stars	Climatic events	Soil erosion	Water flows Water tables	Life cycles Nesting seasons etc.	Ecological succession
Utilitarian	Several	Several	Several	Several	Several	Management units

FIGURE 5 Matrix of indigenous ecological knowledge. Examples of indigenous knowledge that describes relations among events, phenomena, and "natural objects," as well instructs how to manage these resources, are abundant in the literature and are not given in the figure.

from ecosystems. In this context, a predominant use-value economic rationality is adopted, which in practical terms is represented by a multiuse strategy that maximizes the variety of goods produced in order to provide basic household requirements throughout the year (for further details on this strategy, see Toledo, 1990). This main feature accounts for the relatively high self-sufficiency of indigenous households and communities.

Indigenous households tend to carry out a nonspecialized production based on the principle of diversity of resources and practices. This mode of subsistence results in the maximum utilization of all the available landscapes of the surrounding environments, the recycling of materials, energy, and wastes, the diversification of the products obtained from ecosystems, and, especially, the integration of different practices: agriculture, gathering, forest extraction, agroforestry, fishing, hunting, small-scale cattle-raising, and handicrafts. As a result, indigenous subsistence implies the generation of a myriad of products, including food, domestic and work instruments, housing materials, medicines, fuelwoods, fibers, animal forage, and others.

Under the multiuse strategy, indigenous producers manipulate the natural landscape in such a way that two main characteristics are maintained and favored: habitat patchiness and heterogeneity, and biological as well as genetic variation. In the spatial dimension, the landscapes in which indigenous peoples live become a complex landscape mosaic in which agricultural fields, fallow areas, primary and secondary vegetation, household gardens, cattle-raising areas, and water bodies are all segments of the entire production system. This mosaic represents the field upon which indigenous producers, as multiuse strategists, play the game of subsistence through the manipulation of ecological components and processes (including forest succession, life cycles, and movement of materials).

It has been demonstrated that some natural disturbances can increase biodiversity if they increase habitat heterogeneity, reduce the influence of competitively dominant species, or create opportunities for new species to invade the area. On the other hand, the number of species is commonly relatively small in highly disturbed biotic communities, because few populations are able to reestablish themselves before they are reduced by later disturbances. In contrast, a low rate of disturbance provides few opportunities for pioneer species and might allow competitively dominant species to usurp limiting resources. Therefore, biodiversity is often greater at intermediate levels of disturbances than at either lower or higher rates.

The creation of landscape mosaics under the indigenous multiuse strategy in areas originally covered by only one natural community represents a human-originated mechanism that theoretically tends to maintain (and even increase) biodiversity. Several authors have already stressed the importance of the models of low-intensity mosaic usage of the landscape by indigenous peoples and other small-landowner populations for biodiversity conservation.

The same diversified arrangement found in indigenous landscapes tends to be reproduced at the local scale, with multispecies, multistory crops or agroforests favored over monocultures. As a consequence, animal and especially plant genetic resources tend to be maintained in indigenous agricultural fields, aquaculture systems, home-gardens, and agroforests (Gadgil *et al.*, 1993). Polycultural systems managed by indigenous agriculturalists and agroforesters are relatively well known and the recent specialized literature is full of case studies illustrating such designs. Especially notable are the home-gardens and agroforestry systems of the tropical and humid regions of the world, which operate as human-made refuge areas for many species of plants and animals, notably in areas strongly affected by deforestation (Moguel and Toledo, 1999).

At the farm level, it is broadly recognized that crop populations are more diverse in indigenous farming systems than in agricultural areas dominated by industrial agriculture. Therefore, indigenous peoples are regarded as key agents of on-farm preservation of plant genetic resources that are threatened by agricultural modernization (or genetic erosion). The loss of biodiversity is also experienced in farming systems when indigenous cropping polycultural practices are replaced by fossil-fueled monocrops. Indigenous agricultural systems and landscapes are widely acknowledged as designs that preserve not only landraces of crop species, but also semidomesticated and wild crop relatives and even nondomesticated species.

VII. CONSERVING BIODIVERSITY BY EMPOWERING INDIGENOUS PEOPLES

During the past three decades, as the loss of landscapes, habitats, species, and genetic diversity has become an issue of international concern, the protected areas of the world have increased notably in both size and number. However, as protected areas expanded, it became evident that the model of uninhabited national parks originated by the so-called developed nations could not be applied worldwide. Today, there are nearly 10,000 nationally protected areas (parks and other reserves) in more than 160 countries, covering some 650 million of ha, which represents over 5% of the earth's land surface. Many of the areas that have been established as protected areas and many of those that are suitable for future addition to the protected area network are the homelands of indigenous peoples. In Latin America alone, over 80% of protected areas are estimated to have indigenous people living within them. On the other hand, large tracts of the territories under indigenous control, estimated to encompass between 12 and 20% of the earth's surface, are considered to be priority candidates as future reserves. Moreover, some authors, such as J. Alcorn (1994), believe that the bulk of the world's biodiversity is now held within the limits of the indigenous territories in tropical countries.

Given this situation, as well as the evidence offered and discussed in the previous sections, the idea that successful and long-term biodiversity conservation will be impossible without the participation of indigenous communities is gaining recognition in national and international conservation circles. For example, in its latest guidelines, the IUCN (World Conservation Union) Commission on National Parks and Protected Areas (1994) considers that indigenously established "protected territories" can now be recognized as national parks, wilderness areas, protected landscapes, and managed resource protected areas. Furthermore, the international conservation community is beginning to realize that sacred forests, mountains, lakes, rivers, and deserts can also be considered protected areas, as well as managed reefs, lagoons, rivers, and grasslands.

Protected areas based on consultation, comanagement, and even direct management by indigenous peoples are likely to be increasingly important in coming years as the key role of indigenous cultures is gradually recognized and accepted. However, it is important not to idealize indigenous peoples and their resource management strategies and stewardship skills. Some conservationists have been criticized for over-romanticizing indigenous peoples, and in so doing creating a late-twentieth-century version of "the noble savage" (Redford, 1991). Acknowledgment of the positive links between indigenous peoples and biodiversity has been increasingly tempered by the recognition that under certain circumstances (high population densities, market pressures, unsuitable technologies, local disorganization) indigenous peoples can act as disruptive, not as conservationist, actors. In fact, many of the conservation strategies normally applied by indigenous peoples at the local level can be profoundly affected by exogenous phenomena such as economic exploitation, cultural domination, and technological change.

Today, biological diversity and sustainable development are two of the most powerful and central concepts in environmental protection. In recent years, special attention has been paid to supporting the sustainable development of community-based peoples as a key mechanism for the reinforcement of responsible participation of local communities in biodiversity conservation. Sustainable community development can be defined as an endogenous mechanism that allows a local society to take (or retake) control of the processes that affect it. In other words, self-determination and local empowerment, conceived as a "taking of control," have to be the central objectives in all community development.

Given the demonstrated importance of indigenous peoples for biodiversity conservation, it is essential to recognize the necessity of empowering local communities. This will require that indigenous communities be allowed to maintain, reinforce, or assume control of their own territories and natural resources, as well as have sufficient access to relevant information and technology that will assist their resource management. It is important that they hold legally recognized and enforceable rights to lands and waters, which will give the communities both an economic incentive and a legal basis for stewardship. In many countries, national recognition and policy support for existing, community-based property rights systems are crucial. In many Asian and African countries, the return of a measure of control over public lands and resources to local communities is fundamental to slowing biodiversity loss in threatened regions.

Similarly, it is crucial to establish new resource-management partnerships among local communities and provincial and national agencies. Local stewardship, in conjunction with external governmental and nongovernmental institutions and organizations, is perhaps the best way to guarantee the effective protection of landscapes, habitats, species, and genes worldwide, and especially in the biodiversity-rich tropical countries.

VIII. CONCLUDING REMARKS: A BIOCULTURAL AXIOM

The research accumulated in the three last decades by investigators in the fields of conservation biology, linguistics and anthropology of contemporary cultures, ethnobiology, and ethnoecology has converged toward a shared principle: the world's biodiversity will be effectively preserved only by protecting the diversity of human cultures, and vice versa. This principle, which represents a new biocultural axiom, is supported by four main sets of evidence: the geographical overlap between biological richness and linguistic diversity and between indigenous territories and biologically high-value regions (actual and projected protected areas), the recognized importance of indigenous peoples as managers and inhabitants of well-preserved habitats, and the certification of ecologically sustainable behavior among indigenous peoples derived from their premodern belief–knowledge–practices complex.

This bio-cultural axiom, referred to by the late Bernard Nietschmann as the "concept of symbiotic conservation [in which] biological and cultural diversity are mutually dependent and geographically coterminous," constitutes a key principle for conservation theory and application, and epistemologically it is an expression of the new, integrative, interdisciplinary research that is gaining recognition in contemporary science.

See Also the Following Articles

BIODIVERSITY-RICH COUNTRIES • ETHNOBIOLOGY AND ETHNOECOLOGY • HUNTER-GATHERER SOCIETIES • POVERTY AND BIODIVERSITY • SOCIAL AND CULTURAL FACTORS • TRADITIONAL CONSERVATION PRACTICES • TROPICAL ECOSYSTEMS

Bibliography

Alcorn, J. (1993). Indigenous peoples and conservation. *Conservation Biol.* 7, 424–426.

Alcorn, J. (1994). Noble savage or noble state?: Northern myths and southern realities in biodiversity conservation. *Etnoecológica* 3, 7–19.

Burger, J. (1987). *Report from the Frontier: The State of the World's Indigenous Peoples.* Zed Books, London.

Davis, S. H., and A. Wali. (1994). Indigenous land tenure and tropical forest management in Latin America. *Ambio* 23, 207–217.

Durning, A. T. (1993). Supporting indigenous peoples. In *State of the World 1993* (L. Brown, ed.), pp. 80–100. Worldwatch Institute, Washington, D.C.

Food and Agricultural Organization (FAO) of the United Nations (1991). *Statistical Yearbook 1991.* Rome.

Gadgil, M. (1993). Biodiversity and India's degraded lands. *Ambio* 22, 167–172.

Gadgil, M., F. Berkes, and C. Folke. (1993). Indigenous knowledge for biodiversity conservation. *Ambio* 22, 151–156.

Grimes, B. (ed). (1996). *Ethnologue: Languages of the World*, 13th ed. Summer Institute of Linguistics, Dallas.

Hale, K. (1992). On endangered languages and the safeguarding of diversity. *Language* 68, 1–2.

Harmon, D. (1995). The status of the world's languages as reported in the *Ethnologue. Southwest. J. Linguistics* 14, 1–33.

Harmon, D. (1996). Losing species, losing languages: Connections between biological and linguistic diversity. *Southwest. J. Linguistics* 15, 89–108.

Hitchcock, R. K. (1994). International human rights, the environment, and indigenous peoples. In *Endangered Peoples: Indigenous Rights and the Environment*. Colorado Journal of International Environmental Law and Policy, pp. 1–22. University Press of Colorado, Niwot.

Maffi, L. (1999). Language and the environment. In *Cultural and Spiritual Values of Biodiversity* (D. Posey and G. Dutfield, eds.). United Nations Environment Programme, Nairobi, in press.

Maffi, L. (ed.). (2000). *Language, Knowledge and the Environment: The Interdependence of Cultural and Biological Diversity*. Smithsonian Institution Press, Washington, D.C., in press.

Mittermeier, R., and C. Goettsch-Mittermeier. (1997). *Megadiversity: The Biologically Richest Countries of the World*. Conservation International/CEMEX/Sierra Madre, Mexico City.

Moguel, P., and Toledo, V. M. (1999). Biodiversity conservation in traditional coffee systems of Mexico. *Conserv. Biol.* 13, 11–21.

Oldfield, M., and J. Alcorn (eds.). (1991). *Biodiversity: Culture, Conservation and Ecodevelopment*. Westview Press, Boulder, Colorado.

Orlove, B. S., and S. B. Brush. (1996). Anthropology and the conservation of biodiversity. *Annu. Rev. Anthropol.* 25, 329–352.

Redford, K. H. (1991). The ecologically noble savage. *Cultural Survival Quart.* 15, 46–48.

Stevens, S. (ed.). (1997). *Conservation through Cultural Survival: Indigenous Peoples and Protected Areas*. Island Press, Washington, D.C.

Toledo, V. M. (1990). The ecological rationality of peasant production. In *Agroecology and Small-Farm Development* (M. Altieri and S. Hecht, eds.), pp. 51–58. CRC Press, Boca Raton, Florida.

Toledo, V. M. (1992). What is ethnoecology?: Origins, scope and implications of a rising discipline. *Etnoecológica* 1, 5–21.

INSECTICIDE RESISTANCE

Ian Denholm and Greg Devine
IACR—Rothamsted

GLOSSARY

acetylcholinesterase (AChE) The enzyme responsible for breaking down the neurotransmitter acetylcholine (ACh) at nerve synapses, thereby preventing hyperexcitation of cholinergic pathways in the nervous system.

bioassay (biological assay) A laboratory test for evaluating the response of organisms to a toxin and for diagnosing the presence or absence of resistance.

cross-resistance The ability of a single gene or mechanism to confer resistance to more than one toxin.

cytochrome P_{450} monooxygenases A ubiquitous group of enzymes involved in the NAPDH-mediated oxidation and metabolism of a broad range of endogenous and exogenous substrates.

GABA receptor Part of the inhibitory ion channel complex gated by GABA (γ-aminobutyric acid) in postsynaptic nerve membranes.

glutathione S-transferases (GSTs) Enzymes that catalyze the metabolism of a range of substrates following their conjugation with the endogenous tripeptide glutathione.

multiple resistance The occurrence of more than one resistance mechanism in the same individual or pest population.

synergist A chemical used, at sublethal concentrations, to inhibit particular groups of detoxifying enzymes and therefore to implicate the involvement of these enzymes in resistance.

voltage-gated sodium channel A large transmembrane protein that regulates the flow of sodium ions across axonal membranes and mediates the rising phase of action potentials.

THE PHENOMENON OF INSECTICIDE RESISTANCE reflects a genetic adaptation enabling arthropods to survive exposure to otherwise lethal amounts of insecticide. This article reviews the genetic, biochemical, and ecological basis of resistance as well as options for combating its detrimental impact on crop protection and disease management.

Encyclopedia of Biodiversity, Volume 3

I. INTRODUCTION

The diversity of organisms and their genetic variation have been forged by evolution. In evolution, undirected mutations generate unstructured diversity, which is then structured by selection. The time scales over which selection and adaptation occur in eukaryotes are usually too large to observe *in situ* (except in a few cases, such as industrial melanism in some moths and ladybirds), but there is one microevolutionary process for which many of the factors driving selection and adaptation are well understood, a process that is being increasingly studied and manipulated by a large number of biologists. This is the phenomenon of pest resistance to insecticides, and its study can reveal much about how biodiversity originates, at the infraspecific level at least. Similar processes underpin the evolution of herbicide resistance by weeds, fungicide resistance by plant pathogens, and drug resistance by disease-causing microbes.

II. EXTENT OF RESISTANCE

Although a relatively recent phenomenon (resistance to the first synthetic insecticide, DDT, was initially reported in the 1940s), insecticide resistance is now very widespread. Available statistics (Fig. 1) show that reports of resistant arthropod species increased almost exponentially between 1950 and 1980, following the successive introduction of different classes of synthetic insecticides. By 1990, over 500 species were reported to resist chemicals of at least one insecticide class, and many of these resisted several classes simultaneously. Of the resistant species reported in 1990, 88% were insects (class Insecta) and 12% were mites and ticks (class Arachnida, order Acarina). Four insect orders—Coleoptera (beetles), Diptera (true flies), Heteroptera (aphids, bugs, hoppers, and whiteflies), and Lepidoptera (moths)—accounted for 92% of the resistant insect species; the remainder mostly comprised cockroaches, thrips, lice, and fleas.

Although almost all insecticide classes are now affected by resistance, its extent varies greatly between species. In some insects, resistance only extends to a few closely related compounds in a single class; it may be very weak or restricted to a small part of their geographical range. At the other extreme, some widespread pests such as the diamondback moth (*Plutella xylostella*), the Colorado beetle (*Leptinotarsa decemlineata*), the peach–potato aphid (*Myzus persicae*; Fig. 3c), and the cotton whitefly (*Bemisia tabaci*; Fig. 3b) now resist most or all of the insecticides available for their control. The most extensively used insecticide classes—organochlorines, organophosphates, carbamates, and pyrethroids—have generally been the most seriously compromised by resistance, and many principles relat-

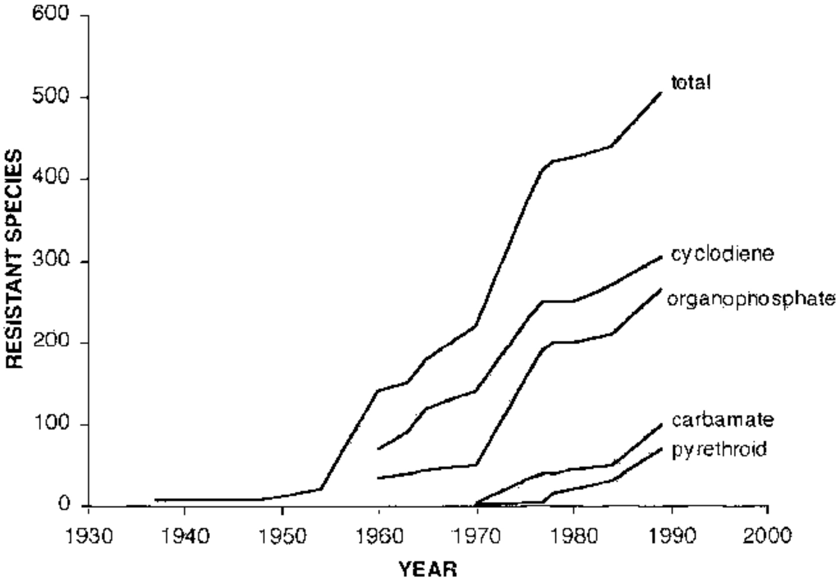


FIGURE 1 Increase in the number of arthropod species reported to resist insecticides over time, in total, and in response to the four most widely used classes of insecticide (courtesy GP Georghiou).

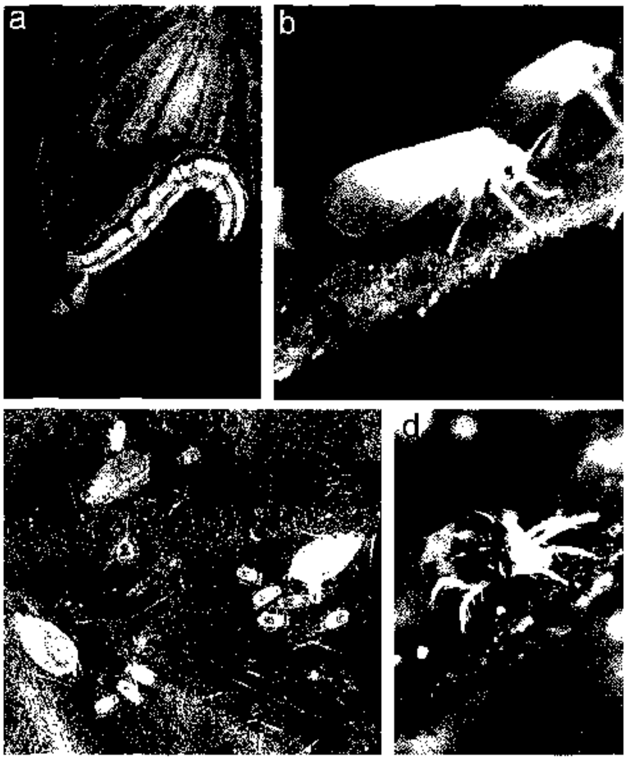


FIGURE 3 Examples of arthropod pests that have developed resistance to insecticides. (a) The boworm, *Helicoverpa armigera*, a major pest of cotton and vegetables in the Old World; (b) the tobacco or cotton whitefly, *Bemisia tabaci*, which threatens a wide range of crops through direct-feeding damage and the transmission of virus disease; (c) the peach–potato aphid, *Myzus persicae*, a major pest of vegetable and ornamental crops in temperate countries; and (d) the two-spotted spider mite, *Tetranychus urticae*, a cosmopolitan pest of fruit, vegetable, and ornamental crops. See also color insert, Volume 1.

ing to the origin and evolution of resistance can be demonstrated solely by reference to these fast-acting neurotoxins. In recent years, however, there has also been a worrying increase in resistance to more novel insecticides, including ones attacking the developmental pathways of arthropods (e.g., benzoylphenylureas), their respiratory processes (e.g., mitochodrial electron transport inhibiting [METI] acaricides), and their digestive systems (e.g., *Bacillus thuringiensis* endotoxins).

III. ORIGINS OF RESISTANCE

Since insecticides are not normally considered mutagenic at field application rates, it is assumed that resistance mutations occur independently of insecticide exposure, and are as likely to occur before an insecticide is introduced as they are during its use in the field. Resistance is therefore a preadaptive phenomenon reflecting the selection of individuals possessing heritable genetic traits that promote their survival or reproduction in environments treated with insecticides. Estimates of mutation rates are as imprecise as they are for other adaptive traits, ranging from 10^{-3} to 10^{-10} depending on the mutational event involved (see below). The role of enzyme induction in resistance has not been demonstrated satisfactorily but is likely to be only slight. Increased tolerance caused by environmental or biological factors such as diet, age, or climate can sometimes be significant but is outside the scope of this article.

IV. MECHANISMS OF RESISTANCE

Some of the most significant recent progress in understanding resistance as an adaptive phenomenon has resulted from the application of molecular biology to resistance research. Depending on the mechanism involved, resistance has been shown to arise through structural alterations of genes encoding target-site proteins or detoxifying enzymes or through processes affecting gene expression (e.g., amplification or altered transcription). Other mechanisms that have been demonstrated or postulated include reduced penetration of insecticides through the insect cuticle, enhanced excretion of insecticides, and behavioral traits enabling pests to reduce or avoid exposure to a toxin. However, the latter are generally considered to be relatively minor in effect or to arise only under very specialized circumstances. Figure 2 is a schematic representation of some of the resistance mechanisms discussed in more detail below.

A. Increased Detoxification of Insecticides

The three major routes of detoxification implicated in resistance are as follows:

Enhanced metabolism of insecticides by cytochrome P_{450} monooxygenases can potentially confer resistance to most chemical classes. However, much of the evidence for this mechanism is indirect, based on the ability of compounds such as piperonyl butoxide (PBO), which are known inhibitors of monooxygenases, to reduce the magnitude of resistance when used as synergists in bioassays. Cases in which enhanced oxidative metabolism has been demonstrated directly through binding or metabolism studies are

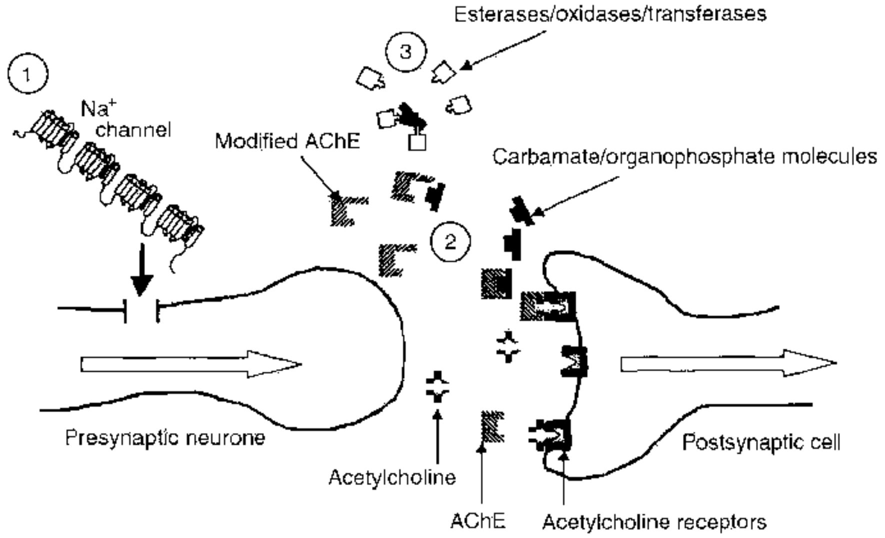


FIGURE 2 Schematic diagram of a nerve synapse showing examples of insecticide resistance mechanisms. (1): Changes in the structure of the sodium channel confer kdr or super-kdr target-site resistance to pyrethroids. (2) Modified acetylcholinesterase is no longer bound by organophosphates and remains available to breakdown acetylcholine molecules after neurotransmission across the synapse. (3) Detoxifying enzymes degrade or sequester insecticides before they reach their targets in the nervous system.

still uncommon. Since it is now apparent that PBO and related molecules can affect other processes in insects, including cuticular penetration and esterase hydrolysis, claims for this type of mechanism based purely on synergism studies should be interpreted with caution.

Enhanced activity of glutathione S-transferases (GSTs) is considered potentially important in resistance to some classes of insecticide, including organophosphates. However, information on the role of GSTs in resistance is still sketchy since this mechanism cannot be diagnosed reliably from bioassays. GSTs, like monooxygenases, exist in numerous molecular forms with distinct properties, making correlations of enzyme activity with resistance very challenging and often ambiguous.

Enhanced hydrolysis or sequestration by esterases (e.g., carboxylesterases) capable of binding to and cleaving carboxylester and phosphotriester bonds undoubtedly plays a significant role in resistance to organophosphates and pyrethroids. Of the three main types of detoxification mechanism, this is by far the best characterized biochemically, and in some cases (e.g., for mosquitoes, blowflies, and the aphid *Myzus persicae*) the esterases involved have been identified, characterized, and sequenced at the molecular level. Resistance due to increased esterase activity can arise through either a qualitative change in an enzyme, improving its hydrolytic capacity, or (as in mosquitoes and aphids) a quantitative change in the titer of a particular enzyme that already exists in susceptible insects.

B. Alterations to Insecticide Target Sites

Resolving target-site modifications that lead to resistance clearly requires some knowledge of the mode of action of insecticides themselves. At present, this information is most advanced for molecules binding to enzymes or receptors in the nervous system of arthropods. Three examples of target site resistance are now well understood:

Organophosphates and carbamates exert their toxicity by inhibiting the enzyme acetylcholinesterase (AChE), thereby impairing the transmission of nerve impulses across cholinergic synapses. Mutant forms of AChE showing reduced inhibition by these insecticides have been demonstrated in several insect and mite species. Biochemical and molecular analyses of insecticide-insensitive AChE have shown that pests may possess several different mutant forms of this enzyme with contrasting insensitivity profiles, thereby conferring distinct patterns of resistance to these two insecticide classes.

Pyrethroids act primarily by binding to and blocking the voltage-gated sodium channel of nerve membranes. A mechanism postulated to reflect insensitivity of this target site was first identified in houseflies (*Musca domestica*) and termed knockdown resistance or kdr. Recently, there has been rapid progress in attributing kdr resistance unequivocally to structural modifications of a sodium channel protein and in locating and identifying the genetic mutations responsible. As a consequence, kdr resistance has now been confirmed in many insect pests. As with insensitive AChE, there can be different forms of kdr resistance (e.g., a more potent "super-kdr" form in houseflies), and this can have important implications for cross-resistance within the pyrethroid class.

GABA receptors are targets for several insecticide classes, including cyclodienes (a subclass of the organochlorines), avermectins, and fipronils. The primary mechanism of resistance to cyclodienes and fipronils in several species involves modification of a particular GABA receptor subunit, resulting in substantial target-site insensitivity to these insecticides.

V. HOMOLOGY OF RESISTANCE GENES

Although there are several possible mechanisms, genetic options for resisting insecticides can also be very limited, especially for mechanisms based on decreased sensitivity of insecticide target sites. The target-site mechanism of cyclodiene resistance has been attributed to the same amino acid substitution (alanine-302 to serine) in GABA receptors of several species of diverse taxonomic origin, including *Drosophila*, several beetles, a mosquito (*Aedes aegypti*), a whitefly (*Bemisia tabaci*), and a cockroach (*Blatella germanica*). Work on the two other principal target-site mechanisms—altered AChE and kdr—has proved more complex due to the occurrence of multiple resistance alleles at the same loci. In the case of kdr, however, there is also evidence for the same amino acid substitution (leucine-1014 to phenylalanine) in a sodium channel protein conferring a "basal" kdr phenotype in a range of species including houseflies, cockroaches, the peach–potato aphid, the diamondback moth, and a mosquito (*Anopheles gambiae*). This phenotype may subsequently be enhanced (to "super-kdr" resistance) by further mutations that also recur between species. Despite the structural complexity of the receptors involved, these parallel mutations imply that the opportunities for insects to modify them to avoid or reduce binding of insecticides, while retaining normal functioning of the nervous system, are very limited indeed.

When susceptible individuals of the sheep blowfly (*Lucilia cuprina*) were exposed to the mutagen ethyl methanesulfonate (EMS) and their progeny screened for resistance to dieldrin (a cyclodiene), surviving insects not only exhibited a GABA receptor based mechanism analogous to that found in nature but also exhibited an identical alanine to serine amino acid substitution in the channel gene. Similarly, mutagenesis followed by screening with diazinon (an organophosphate) led to the recovery of a resistance mechanism showing identical toxicological, biochemical, genetic, and molecular properties to one that had previously evolved to diazinon under field conditions. These findings again demonstrate tight evolutionary constraints on the number of viable resistance mutations, even in the laboratory, where mutations conferring deleterious effects on overall fitness might be expected to survive better than in the open field. Interestingly, they also highlight the potential of using mutagenesis to predict likely resistance mechanisms to novel insecticides well in advance of them appearing in field populations and to tailor resistance management recommendations accordingly.

In other cases, different types of resistance to the same toxin exist and can account for differences in the toxicological and genetic basis of resistance between species or between different geographical populations of the same species. Insecticidal proteins from the soil bacterium *Bacillus thuringiensis* (Bt) are becoming increasingly important in pest management, especially in relation to insect-tolerant transgenic crops (see below). To date, the only species to have evolved Bt resistance in the field is the diamondback moth, *Plutella xylostella*. The majority of Bt-resistant populations examined have exhibited very similar characteristics, including a very consistent pattern of cross-resistance to different Bt toxins and recessive inheritance. However, there are also strains of *P. xylostella* in which the breadth and inheritance of Bt resistance differ markedly from this "mode," implying the existence of distinct resistance genes and/or mechanisms.

VI. HOW OFTEN DO RESISTANCE GENES ARISE?

The recurrence of specific resistance mutations within and between taxa begs another question of fundamental significance to the origins of biodiversity: Have such mutations arisen repeatedly within the same species, or appeared on only a limited number of occasions and subsequently spread through migration and/or human agency? This question is proving amenable to investigation by sequencing not only the resistance genes themselves but also flanking regions and introns, which would be expected to vary between alleles that have arisen independently. In the mosquito, *Culex pipiens*, organophosphate resistance is primarily conferred by allozymes at two closely linked loci (esterases A and B) coding for insecticide-detoxifying carboxylesterases. Overproduced allozymes (resulting from amplification of A or B genes) tend to recur in geographically disjunct areas. This situation could be explained by recurrent mutation generating each amplification event *de novo* or by a nonrecurrent mutation that has subsequently spread within and between populations. Restriction mapping of DNA around the esterase genes points to the latter explanation, with large-scale gene flow (even between continents) most likely attributable to passive migration of mosquitoes on ships and/or airplanes. It is notable that the recent appearance of a new resistance allele in southern France is known to have originated in the vicinity of the international airport and seaport at Marseilles.

Organophosphate resistance in the aphid *Myzus persicae* is also attributable to the amplification of a gene encoding an insecticide-detoxifying carboxylesterase. Despite the often widespread dispersion of these amplified genes in the aphid genome, restriction analyses have indicated that all copies are in the same immediate genetic background. This suggests that amplification occurred only once, with the amplified DNA subsequently being moved intact around the genome through chromosomal rearrangements, or perhaps mediated by transposable elements. Similarities in the position and structure of these genes in aphids of diverse geographic origin reinforce the likelihood of a single amplification event that has subsequently become widely dispersed around the world.

There is also molecular evidence for some resistance genes having several independent origins in the same species (e.g., for target-site resistance to cyclodienes in the red flour beetle, *Tribolium castaneum*). However, results for mosquitoes and aphids highlight the potential for large-scale inadvertent movement of resistant insects between countries or even continents. For crop pests, these risks are particularly acute due to the increasing international trade in edible and ornamental plants, many of which have been treated with insecticides at their point of origin. In such cases, growers at the receiving end of the trade network face a dual threat: (i) the establishment of new pest species or more aggressive biotypes of existing ones and (ii) the possibility that such pests are already strongly resistant to compounds that might otherwise be used to suppress them or to eradicate them entirely.

VII. CROSS-RESISTANCE AND MULTIPLE RESISTANCE

Arthropods seldom if ever resist just one toxin. Most commonly, they exhibit differing levels of resistance to a range of related and unrelated insecticides. In its strictest sense, the term *cross-resistance* refers to the ability of a single mechanism to confer resistance to several insecticides simultaneously. A more complex situation is that of *multiple resistance*, reflecting the coexistence of two or more resistance mechanisms, each with their own specific cross-resistance characteristics. Disentangling cross-resistance from multiple resistance, even at the phenotypic level, is one of the most challenging aspects of resistance research. However, a knowledge of the mechanisms involved is often essential in order to develop resistance management recommendations based, for example, on the alternation of insecticides to avoid continuous selection for the same resistance gene or mechanism.

Unfortunately, cross-resistance patterns are inherently difficult to predict in advance, since mechanisms based on both increased detoxification and altered target sites can differ substantially in their specificity. The most commonly encountered patterns of cross-resistance tend to be limited to compounds within the same chemical class (equivalent to the term *side resistance* as used by parasitologists). However, even these patterns can be very idiosyncratic. For example, organophosphate resistance based on increased detoxification or target-site alteration can be broad-ranging across this group or highly specific to a few chemicals with particular structural similarities. The breadth of target-site resistance to pyrethroids in houseflies is also dependent on the resistance allele present. The kdr allele itself affects almost all compounds in this class to a similar extent (ca. 10-fold resistance), whereas resistance due to the more potent super-kdr allele is highly dependent on the alcohol moiety of pyrethroid molecules, ranging from ca. 10-fold to virtual immunity. Cross-resistance between insecticide classes is even harder to anticipate, especially for broad-spectrum detoxification systems whose specificity depends not on insecticides having the same mode of action but on the occurrence of common structural features that bind with detoxifying enzymes.

Empirical approaches for distinguishing between cross-resistance and multiple resistance include (i) repeated back-crossing of resistant populations to fully susceptible ones to establish whether resistance to one chemical cosegregates consistently with resistance to another and (ii) reciprocal selection experiments whereby populations selected for resistance to one chemical are examined for a correlated change in response to another. If available, biochemical or molecular diagnostics for specific resistance genes can assist considerably with tracking the outcome of genetic crosses or with assigning cross-resistance patterns to particular mechanisms.

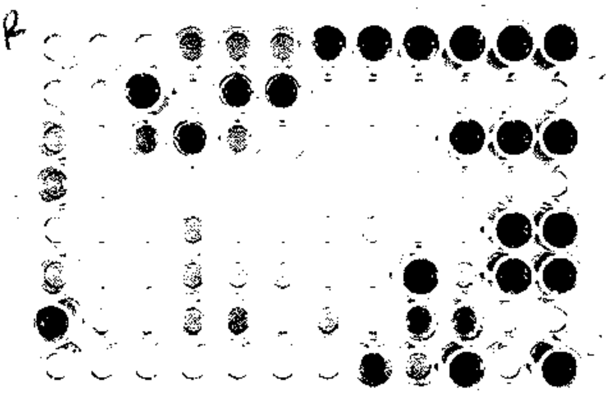


FIGURE 4 Results of a biochemical immunoassay to determine levels of an insecticide-detoxifying esterase in individual peach–potato aphids. The top row shows results for three insects from each of four resistance categories, with resistance increasing from left to right. The remaining insects were drawn randomly from a field sample.

VIII. DIAGNOSIS OF RESISTANCE

Although a large number of laboratory bioassay methods have been developed for detecting and characterizing resistance, most of these are limited to defining phenotypes and provide little or no information on the underlying genes or mechanisms. Thus, although bioassays remain the mainstay of most large-scale resistance monitoring programs, much attention is being paid to developing more incisive techniques that not only offer greater precision and throughput but also diagnose the type of mechanism(s) present and, whenever possible, the genotypes of resistant insects. A variety of approaches are being adopted for this purpose, including electrophoretic or immunological detection of resistance-causing enzymes, kinetic and end-point assays for quantifying the activity of enzymes or their inhibition by insecticides, and DNA-based diagnostics for mutant resistance alleles. The sensitivity that these techniques can provide is exemplified well by work on the aphid *Myzus persicae*, which in northern Europe possesses at least three coexisting resistance mechanisms: (i) an overproduced carboxylesterase conferring resistance to organophosphates, (ii) an altered AChE conferring resistance to certain carbamates, and (iii) target-site (kdr) resistance to pyrethroids. These mechanisms collectively provide strong resistance to virtually all available aphicides. Fortunately, it is now possible to diagnose all three in individual aphids using an immunoassay for the overproduced esterase, a kinetic microplate assay for the mutant AChE, and a PCR-based diagnostic for the kdr allele (Fig. 4). The combined use of these techniques against field populations provides up-to-date information on the incidence of the mechanisms and is used to alert growers to potential control problems.

IX. SELECTION OF RESISTANCE GENES

The rate at which resistance genes are selected reflects the combined influence of numerous biotic and abiotic factors. Resistance offers several advantages for research to resolve these factors and their interactions. First, the selecting agent (exposure to insecticides) is well understood; it can usually be carefully documented using treatment histories or manipulated to investigate its effect on selection rates. Second, the selective advantages conferred by resistance genes are often very large, leading to substantial changes in genetic composition over a measurable time frame. Third, most of the major mechanisms of insecticide resistance, unlike those for many stress-related adaptations, are controlled by single genes of major effect (monogenic) rather than many genes, each of small effect (polygenic). This renders resistance readily amenable to analysis within the conventional theoretical framework of ecological genetics. Finally, the frequent availability of bioassays for quantifying the frequency of resistance phenotypes, or even *in vitro* assays for specific genotypes, enables accurate documentation of responses to selection applied in population cages in the laboratory or under open field conditions.

Factors determining the selection (and hence risk) of resistance to insecticides can, for convenience, be classified into genetic or ecological ones relating to the intrinsic properties of pests and resistance mechanisms and operational ones relating to the chemical itself and how it is applied. Some of the most important factors apparent from the large body of experimental and theoretical research on resistance selection are summarized below.

A. Genetic Influences

In order for resistance to evolve, resistance genes must confer a selective advantage over their susceptible counterparts. One of the primary challenges for describing resistance is therefore to estimate the relative fitness of different genotypes under exposure to insecticides. There are different ways of achieving this, the most direct being to release individuals of known susceptible and resistance genotypes into insecticide-treated environments and to monitor their survival. This has been attempted for a variety of pest species and has identified many, often subtle, influences on how resistance genes are expressed in the field. The dominance of resistance genes, which exerts a major influence on selection rates, is a case in point. In laboratory bioassays evaluating the relative survival of susceptible homozygotes (SS), heterozygotes (RS), and resistance homozygotes (RR) over several insecticide concentrations, dominance can be measured precisely, with RS individuals usually responding in an intermediate manner. In the field, dominance is a changeable phenomenon, depending on the concentration of insecticide applied and its uniformity over space and time. Even when the initial concentration is sufficient to kill RS individuals (rendering resistance effectively recessive), the weathering or decay of residues may result in this genotype showing increased survival and resistance becoming functionally dominant in expression. When resistance genes are still rare, and hence mainly present in heterozygous condition, this can have a profound effect in accelerating the selection of resistance genes to economically damaging frequencies.

The diverse mating systems of insects also influence the rate at which resistance evolves. Although most research has focused on outcrossing diploid species (typified by members of the Lepidoptera, Coleoptera, and Diptera), systems based on haplodiploidy and parthenogenesis also occur among key agricultural pests. In haplodiploid systems, males are produced uniparentally from unfertilized, haploid eggs, and females are produced biparentally from fertilized, diploid eggs. The primary consequence of this (exemplified by whiteflies, spider mites (Fig. 3d), and phytophagous thrips) is that resistance genes are exposed to selection from the outset in haploid, hemizygous males, irrespective of intrinsic dominance or recessiveness. Whether a resistance gene is dominant, semidominant, or recessive, resistance can develop at a similar rate under haplodiploidy, whereas recessiveness can cause significant delays (initially at least) in diploid populations.

Most species of aphid undergo periods of parthenogenesis, promoting the selection of clones with the highest levels of resistance and/or the most damaging combination of resistance mechanisms. However, in holocyclic populations (ones that alternate between sexual and asexual reproduction), this effect is at least partially countered by genetic recombination and the subsequent reassortment of mechanisms during sexual reproduction. In fully anholocyclic (asexual) populations, such as those of *Myzus persicae* in northern Europe, the influence of parthenogenesis is much more severe and has led to strong and persistent associations between resistance mechanisms within clonal lineages exposed to a succession of different selecting agents.

B. Ecological Influences

Several aspects of pest ecology, including the dynamics, phenology, and dispersal capabilities of pest organisms, act as primary determinants of resistance development. However, their influence on selection rates can be unpredictable without a sound knowledge of how they interact with patterns of insecticide use. As an example, movement of pests between untreated and treated parts of their range may delay the evolution of resistance, due to the diluting effect of susceptible immigrants. Conversely, large-scale movement can, as described above, also accelerate the spread of resistance by transferring resistance alleles between localities.

For highly polyphagous crop pests, interactions between pest ecology and insecticide treatments play a particularly critical role in determining selection pressures. Key factors to be considered are the seasonality and relative abundance of treated and untreated plant hosts and patterns of migration between hosts at different times of the year. A good example relates to the two major bollworm species (Lepidoptera: Noctuidae) attacking cotton in Australia. Only *Helicoverpa armigera* (Fig. 3a) has developed strong resistance; *H. punctigera*, despite being an equally important cotton pest, has remained susceptible to all insecticide classes. The most likely explanation is that *H. punctigera* occurs in greater abundance on a larger range of unsprayed hosts than *H. armigera*, thereby precluding a significant increase in resistance on treated crops.

However, polyphagy can sometimes be deceptive. In the cotton/vegetable/melon production systems of the southwestern United States, the highly polyphagous whitefly *Bemisia tabaci* has become a devastating pest and a primary target of insecticide sprays. The consequences for resistance have varied substantially on a regional basis. Resistance problems have proved much more severe and persistent on cotton in south-central Arizona than in the extreme southwest of the state and the adjacent Imperial Valley of California. This appears attributable to the higher proportion of unsprayed hosts, especially alfalfa, acting as a buffer to resistance in the latter areas and preventing any directional increase in the severity of resistance over successive seasons. In south-central Arizona, these untreated refuges are much less abundant during the time that cotton is treated with insecticides. Thus, a large proportion of the local whitefly population is forced through a selection "bottleneck" on cotton and exposed to intense selection for resistance, despite an abundance of alternative hosts at other times of the year.

Enclosed environments such as greenhouses and glasshouses, which restrict migration and escape from insecticide exposure under climatic regimes favoring rapid and continuous population growth, provide ideal ecological conditions for selecting resistance genes. Very low or zero damage tolerance thresholds for high-value ornamental or vegetable produce accentuate the problem by promoting overfrequent spraying and hence intensify selection for resistance. Over the years, these environments have proved potent sources of novel resistance mechanisms for a diverse range of control agents and have presented a particular challenge to attempts at resistance management.

C. Operational Influences

Although closely linked to aspects of pest genetics and ecology, operational factors are best distinguished as ones which, in principle at least, are at man's discretion and can be manipulated to influence selection rates. Factors exerting a major influence in this respect include the rate, method, and frequency of applications, their biological persistence, and whether insecticides are used singly or as mixtures of active ingredients.

Equating operational factors with selection is often difficult, since without a detailed knowledge of the resistance mechanisms present it is impossible to test many of the assumptions on which genetic models of resistance are based. Anticipating the selection pressure imposed by a particular application dose of insecticide is a case in point. If resistance alleles are present, the only entirely nonselecting doses will be ones sufficiently high to overpower all individuals, regardless of their genetic composition, or ones sufficiently low to kill no insects at all. The latter is obviously a trivial option. Prospects of achieving the former depend critically on the potency and dominance of resistance genes present. A pragmatic solution to this dilemma is to set application doses as far above the tolerance range of SS individuals as economic and environmental constraints permit, in the hope that at least RS genotypes will be effectively controlled. Even this approach can backfire badly if resistance turns out to be more common than suspected (resulting in the presence of RR homozygotes) or resistance alleles exhibit an unexpectedly high degree of dominance. Unless a high proportion of insects escape exposure altogether, the consequence could then be to select very rapidly and effectively for homozygous resistant populations.

In practice, concerns over optimizing dose rates to avoid resistance are secondary to ones regarding the application process itself. Delivery systems and/or habitats promoting uneven or inadequate coverage will generally be more prone to selecting for resistance, since pests are more likely to encounter exposure conditions under which selection is most intense. This was elegantly demonstrated through experiments assessing the relative survival of cyclodiene-susceptible and -resistant phenotypes of the coffee berry borer (*Hypothemus hampei*) in coffee plantations treated with this chemical in New Caledonia. The practice of spraying plantations from roadsides with vehicle-mounted mistblowers generated gradients in the concentration of endosulfan that resulted in different selection pressures in different parts of each field. Similarly, underdosing with the fumigant phosphine in inadequately sealed grainstores has been implicated as a primary cause of resistance to this chemical in a range of stored product pests.

The timing of insecticide applications relative to the life cycle of a pest can also be an important determinant of resistance. A good example relates to the selection of pyrethroid resistance in the cotton bollworm, *Helicoverpa armigera*, in Australia. On cotton foliage freshly treated with the recommended field dose, pyrethroids killed larvae up to 3–4 days old irrespective of whether they were resistant or not by laboratory criteria. Since the sensitivity of larvae of all genotypes to pyrethroids was found to decline with increasing larval size, the greatest discrimination between susceptible and resistant phenotypes occurred only when larvae achieved a threshold age. Targeting of insecticides against newly hatched larvae, as is generally advocated for bollworm control, not only increases the likelihood of contacting larvae at the most exposed stage in their development but also offers the greatest prospect of retarding resistance by overpowering its expression.

In practice, persistent insecticides are often essential to ensure an acceptable period of control, especially when contending with disease vectors or continued invasion of crop pests from alternative host plants. However, persistent applications can accentuate resistance development by exposing a larger number of individuals to the selecting agent. Another problem is that residues of persistent insecticides decay or become diluted through plant growth, so that resistant insects may survive more effectively than they did at the time of application. Empirical studies with a range of pests including mosquitoes, bollworms, and blowflies have demonstrated that aged deposits discriminate more readily between genotypes or phenotypes than ones freshly applied.

In theory, the coapplication of two or more unrelated chemicals as insecticide mixtures offers substantial benefits for delaying the selection of resistance. The underlying principle is one of "redundant killing," whereby any individuals already resistant to one insecticide are killed by simultaneous exposure to another, and vice versa. However, achieving this objective requires not only that each type of resistance is still rare but also that the ingredients confer mutual protection throughout the effective life of an application. Failure to ensure that they exhibit similar biological persistence may lead to one compound exerting greater selection pressure than the other, thereby accelerating the selection of doubly resistant phenotypes. Two potentially conflicting challenges of choosing ideal mixture partners are therefore (i) to ensure maximum similarity in efficacy and persistence against the target pest(s) and (ii) to ensure maximum dissimilarity in chemical structure and mode of action to minimize the likelihood of cross-resistance. Difficulties with identifying candidate molecules that meet all these criteria have greatly limited the use of mixtures for combating resistance to conventional insecticides, although they have considerable appeal for sustaining the effectiveness of insect-tolerant transgenic crops (see below).

D. Fitness of Resistant Individuals

Despite the advantages they confer under exposure to insecticides, it is often assumed that resistance genes also confer physiological costs that could lead to counterselection when insecticides are not applied. Some of the best examples of such fitness costs come from studies conducted under harsh or stressful environmental conditions, when even slight differences in relative fitness are likely to have major consequences for the survival of genotypes. For example, resistant strains of bollworms, blowflies, and aphids have all been demonstrated to overwinter less successfully than their susceptible counterparts. Possible explanations for these fitness differentials include the reduced viability of certain life stages, a slower reproductive rate rendering resistant insects more vulnerable to adverse climatic conditions or to predation, or a reduced ability to respond to environmental cues promoting survival. In the aphid *Myzus persicae*, resistant individuals are less inclined to move from senescing to younger leaves and are therefore more vulnerable to isolation and starvation after leaf abscission. These costs have been implicated in the decline in frequency of resistant genotypes in the absence of insecticide pressure, leading to a balancing polymorphism for resistance rather than a consistent accumulation of resistance from one season to another.

Fitness costs associated with resistance can be difficult to demonstrate experimentally, since deleterious effects may only be expressed under particular environmental conditions or conferred by other genes closely linked to the resistance locus. The most convincing examples are ones in which costs have been found consistently in resistant populations of diverse geographic origin or have persisted after several generations of back-crossing to susceptible insects in order to exclude linkage effects. The potential for fitness drawbacks to be overcome by a process of coadaptation, i.e., the integration of resistance genes with other "modifier" loci that ameliorate fitness costs, has also proved challenging to demonstrate. In a few cases, however, repeated back-crossing of resistant insects to susceptible ones has led to seemingly fit resistance phenotypes acquiring a fitness penalty, apparently due to the uncoupling of resistance genes from modifier loci.

X. COMBATING INSECTICIDE RESISTANCE

In most studies of evolution, the primary challenge is to identify selective forces and to interpret their effects on the genetic composition of individuals and populations. With insecticide resistance, it is also necessary to intervene in the evolutionary process and find ways of reducing its deleterious impact on pest management. Failure to do so in the past has had many severe consequences, including the economic failure of cropping systems, the resurgence of insect-transmitted pathogens, and damage to the environment by way of increased insecticide applications.

The concept of insecticide resistance management (IRM) aims to address these concerns through the development of control strategies for overcoming resistance to currently used compounds, or preventing its appearance in the first place. Although drawing extensively on the theoretical and empirical framework that evolutionary biology provides, IRM strategies must also contend with several practical, economic, and political constraints on the choice of possible management tactics and the precision with which they can be applied. The most important of these are as follows:

The properties of any resistance genes present will often be unknown, and knowledge of pest ecology may still be rudimentary.

It is often necessary to contend with a whole pest complex rather than just a single pest species.

There will often be a very limited number of insecticides available for use in management strategies.

For highly mobile pests at least, countermeasures may need to be standardized and synchronized over large areas, sometimes whole countries.

Resistance is a dynamic phenomenon; i.e., any mechanisms already known to exist may change over time. Continued monitoring is vital to determine whether management recommendations remain valid or need to be revised in light of changing circumstances or new knowledge gained.

To promote compliance with management strategies, the countermeasure adopted should be as unambiguous, rational, and simple as possible.

A strategy first implemented on Australian cotton in 1983 to contend with the bollworm, *Helicoverpa armigera*, illustrates many features of large-scale attempts at resistance management. It was introduced in response to unexpected, but still localized, outbreaks of pyrethroid resistance in *H. armigera* and was based primarily on the concept of insecticide rotation. The threat of pyrethroid resistance was countered by restricting these chemicals to a maximum of three sprays within a prescribed time period coincident with peak bollworm damage. Farmers were required to use alternative insecticide classes at other stages of the cropping season, in order to diversify the selection pressures being applied.

Compliance with this strategy was excellent, and initially it had the desired effect of preventing a systematic increase in the frequency of pyrethroid-resistant phenotypes. Additional recommendations resulting from work on the ecological genetics of pyrethroid resistance, including the targeting of insecticides against newly hatched larvae (when even resistant insects can be killed) and the plowing-in of cotton stubble to destroy resistant pupae overwintering in the soil, undoubtedly contributed to this success. Unfortunately, the restrictions placed on pyrethroid use were inadequate to prevent a gradual, long-term buildup of pyrethroid resistance. As a result, pyrethroids are no longer considered reliable control agents for *H. armigera*, although they remain highly effective against a coexisting species, *H. punctigera*. The strategy has therefore been revised extensively to place greater emphasis on distinguishing between the two *Helicoverpa* species and on the strategic use of nonpyrethroids against *H. armigera*. Transgenic cotton plants expressing *Bacillus thuringiensis* (Bt) toxins have since been released commercially in Australia. Tactics for deploying these without selecting rapidly for Bt resistance are therefore being investigated as a matter of urgency (see below).

Another strategy incorporating a wide range of chemical and nonchemical countermeasures was introduced on Israeli cotton in 1987. This had the primary objective of conserving the effectiveness of insecticides against the whitefly, *Bemisia tabaci*. Under recommendations coordinated by the Israeli Cotton Board, important new whitefly insecticides are restricted to a single application per season within an alternation strategy optimized to contend with the entire cotton pest complex and to exploit biological control agents to the greatest extent possible. One major achievement of this strategy has been a dramatic reduction in the number of insecticide applications against the whole range of cotton pests, but especially against *B. tabaci*. Sprays against whiteflies now average less than two per growing season compared with over 14 per season in 1986. Most importantly of all, the strategy has generated an ideal environment for releasing additional new insecticides onto cotton and for managing them effectively from the outset.

XI. RESISTANCE IN NONPEST SPECIES

Compared to its prevalence in arthropod pests, insecticide resistance is still relatively rare among nonpest species including beneficial organisms. However, it has been well documented in a few species of hymenopteran parasitoids and predatory mites, some of which are being exploited in integrated pest management (IPM) systems. Its rarity among beneficial organisms is probably due in part to difficulties in locating hosts and prey (and hence surviving) under exposure to insecticides. The likelihood of resistance developing in beneficial arthropods may be increased if the insects they depend on as prey are already resistant, although this requires further research. It is also likely that, in comparison with herbivorous species, the enzyme systems of predators and parasites are less well adapted to detoxify xenobiotics.

The propensity for beneficial insects to evolve resistance obviously depends on the degree of selectable variation within their populations. Although laboratory selection has been used to enhance low levels of pesticide tolerance found in field populations, there is a concern that such selection applied at artificially low doses will promote polygenic traits that could fragment and dissipate if released into natural populations. However, when substantial resistance has evolved naturally in the field, its mechanisms have tended to be similar to ones found in pest species. Organophosphate-resistant strains of the green lacewing, *Chrysopa scelestes*, have been shown to exhibit increased activity of acetylcholinesterase (AChE) compared to susceptible insects. A carboxylesterase enzyme, very similar in amino acid sequence to that conferring organophosphate resistance in the aphid, *Myzus persicae*, has been cloned and sequenced from a malathion-resistant strain of the parasitic wasp *Anisopteromalus calandrae*.

Since the development of resistance is dependent on the ecology of systems in which it appears, interactions between beneficial and pest species will greatly affect the epidemiology and dynamics of resistance in both. For example, some parasitoids of stored-grain beetles are resistant to insecticides, and it is thought that this adaptation has been encouraged by the fact that their parasitic larvae are sheltered from insecticides by the grain kernels inhabited by their hosts. This is thought to protect a substantial part of the insect life cycle from insecticide selection and ensures that relatively small shifts in insecticide tolerance by the parasitoid afford significant protection against the decreased insecticide doses that do penetrate their defenses.

Natural enemies may contribute to retarding resistance in pest species by exerting sufficient control to decrease the number of insecticide treatments required. Conversely, there are ways in which natural enemies could promote the adaptation of pests to insecticides. For example, the selection pressure for resistance would be increased if weaker, sublethally affected individuals were more easily preyed upon or parasitized than their fully resistant and therefore unaffected counterparts.

XII. THE SPECIAL CASE OF TRANSGENIC PLANTS

A new development in crop protection with an important bearing on resistance is the release of crop plants genetically engineered to express genes for insecticidal toxins derived from the microbe *Bacillus thuringiensis* (Bt). Bt cotton and/or corn is already being grown commercially on a large scale in the United States, Canada, Australia, Mexico, China, and South Africa. In 1998, the total area worldwide planted with Bt crops was estimated to exceed 12 million ha. Existing toxin genes in Bt cotton and corn are active specifically against certain key lepidopteran pests (especially bollworms and corn borers); another engineered into potatoes provides protection against the Colorado beetle, *Leptinotarsa decemlineata*.

Aside from their commercial prospects, insect-tolerant transgenic crops offer numerous potential benefits to agriculture. The incorporation of Bt genes into crops offers constitutive expression of toxins in plant tissues throughout a growing season. This could reduce dramatically the use of conventional broad-spectrum insecticides against insect pests as well as remove the dependence of pest control on extrinsic factors, including climate and the efficiency of traditional application methods. However, this high and persistent level of expression also introduces a considerable risk of pests adapting rapidly to resist genetically engineered toxins. To date, there are no substantiated reports of resistance selected directly by exposure to commercial transgenic crops, but resistance to conventional Bt sprays (selected in either the laboratory or the field) has been reported in more than a dozen species of insect. Research into the causes and inheritance of such resistance is providing valuable insights into the threats facing Bt plants and the efficacy of possible countermeasures.

Tactics proposed for sustaining the effectiveness of Bt plants have many parallels with ones considered for managing resistance to conventional insecticides. However, they are more limited in scope due to the long persistence and constitutive expression of engineered toxins and the limited diversity of transgenes currently available. Indeed, for existing "single-gene" plants, the only prudent and readily implementable tactic is to ensure that substantial numbers of pests survive in nontransgenic "refuges," composed either of the crop itself or of alternative host plants. In the longer term, stacking (pyramiding) of two or more genes in the same cultivar, or possibly rotations of cultivars expressing different single toxins, are potentially more durable options for resistance management. Whatever measures are adopted, it is essential that Bt plants (and their successors expressing other transgenes) are exploited as components of multitactic strategies rather than as a panacea for existing pest management problems, including those arising from the development of resistance to conventional insecticides.

XIII. CONCLUDING REMARKS

Over the past 20 years, few areas of entomology have advanced as rapidly or received such widespread attention as that of insecticide resistance. Research on this topic has provided invaluable insights into the origin and nature of adaptations, and these are in turn proving of much broader significance for understanding genetic responses to manmade change in the environment. In many respects the continuing battle against resistance is analogous to an evolutionary "arms race," in this case pitting human ingenuity in discovering new toxins against the adaptive capacity of pest species. Debates as to who will eventually win this race are of secondary importance to the realization that for many species the race is probably unnecessary. A wider adoption of resistance management practices, especially through greater exploitation of nonchemical measures, would assist with reducing both the economic impact of resistance and the deleterious effects of many existing pest and disease management strategies on biodiversity in general.

See Also the Following Articles

DIFFERENTIATION • ECOLOGICAL GENETICS • GENES, DESCRIPTION OF • INSECTS, OVERVIEW • PESTICIDES, USE AND EFFECTS OF

Bibliography

Denholm, I., and Rowland, M. W. (1992). Tactics for managing pesticide resistance in arthropods: Theory and practice. *Annu. Rev. Entomol.* 37, 91–112.

Denholm, I., Cahill, M., Dennehy, T. J., and Horowitz, A. R. (1998). Challenges with managing insecticide resistance in agricultural pests, exemplified by the whitefly *Bemisia tabaci*. *Philos. Trans. R. Soc. London, Ser. B: Biol. Sci.* 353, 1757–1767.

Devonshire, A. L., Field, L. M., Foster, S. P., Moores, G. D., Williamson, M. S., and Blackman, R. L. (1998). The evolution of insecticide resistance in the peach–potato aphid, *Myzus persicae*. *Philos. Trans. R. Soc. London, Ser. B: Biol. Sci.* 353, 1677–1684.

ffrench-Constant, R. H., Pittendrigh, B., Vaughan, A., and Anthony, N. (1998). Why are there so few resistance-associated mutations in insecticide target genes? *Philos. Trans. R. Soc. London, Ser. B: Biol. Sci.*, 353, 1685–1693.

Forrester, N. W., Cahill, M., Bird, L. J., and Layland, J. K. (1993). Management of pyrethroid and endosulfan resistance in *Helicoverpa armigera* (Lepidoptera, Noctuidae) in Australia. *Bull. Entomol. Res., Suppl. 1.*

Horowitz, A. R., Forer, G., and Ishaaya, I. (1994). Managing resistance in *Bemisia tabaci* in Israel with emphasis on cotton. *Pest. Sci.* 42, 113–122.

Martinez-Torres, D., Devonshire, A. L., and Williamson, M. S. (1997). Molecular studies of knockdown resistance to pyrethroids: Cloning of domain II sodium channel gene sequences from insects. *Pest. Sci.* 51, 265–270.

McKenzie, J. A. (1996). *Ecological and Evolutionary Aspects of Insecticide Resistance*. R. G. Landes, Austin, TX.

Raymond, M., Chevillon, C., Guillemaud, T., Lenormand, T., and Pasteur, N. (1998). An overview of the evolution of overproduced esterases in the mosquito *Culex pipiens*. *Philos. Trans. R. Soc. London, Ser. B: Biol. Sci.* 353, 1707–1711.

Roush, R. T. (1989). Designing resistance management programs: How can you choose? *Pest. Sci.* 26, 423–441.

Roush, R. T. (1997). Bt-transgenic crops: Just another pretty insecticide or a chance for a new start in resistance management? *Pest. Sci.* 51, 328–334.

Soderlund, D. M., and Bloomquist, J. R. (1990). Molecular mechanisms of insecticide resistance. In *Pesticide Resistance in Arthropods* (R. T. Roush and B. E. Tabashnik, Eds.), pp. 58–96. Chapman & Hall, London.

Tabashnik, B. E. (1994). Evolution of resistance to *Bacillus thuringiensis*. *Annu. Rev. Entomol.* 39, 47–79.

Tabashnik, B. E., Liu, Y. B., Malvar, T., Heckel, D. G., Masson, L., and Ferre, J. (1998). Insect resistance to *Bacillus thuringiensis*: Uniform or diverse? *Philos. Trans. R. Soc. London, Ser. B: Biol. Sci.* 353, 1751–1756.

INSECTS, OVERVIEW

Brian V. Brown
Natural History Museum of Los Angeles County

GLOSSARY

hexapod Group including insects and their primitive relatives.

instar Stage between molts of immature insects.

pro-, meso-, and metathorax First, second, and third segments of the thorax.

THE INSECTS are the most species-rich group of organisms known, with the most diverse natural histories of any animals. The most successful insect groups have wings for flight, can fold these wings back over the body, and have development that is divided into four discrete stages: egg, larva, pupa, and adult.

I. INTRODUCTION

The Class Insecta, or the slightly larger Superclass Hexapoda (which additionally includes the orders Collembola, Protura and Diplura), is the world's most species-rich group of organisms, with about 1 million described species. They are found nearly everywhere on earth, including the terrestrial, aquatic, and, to a much lesser extent, marine ecosystems. They are most diverse in tropical forests, where the undescribed fauna has been estimated to comprise 5, 10, 30, or even 50 million species. Hexapods are a well-established monophyletic group, based on the presence of three major body divisions—head, thorax, abdomen—and a single pair of locomotory appendages on each thoracic segment. Some primitive insects have retained appendages on the abdominal segments, but these are much smaller and less functional than those on the thorax. Most of the more derived groups of insects also have wings, usually one pair on each of the mesothoracic and metathoracic segments, but these have been variously lost or modified in some groups, especially the Diptera (flies).

The fossil record of hexapods extends back to the earliest record of terrestrial life, with Collembola and lower insects recorded from the lower Devonian, almost 400 million years ago, and possibly even earlier traces from the Silurian. Insects have been prominent members of the fossil record ever since, with most prominent major groups having been preserved from late Paleozoic or early Mesozoic formations (200–250 million years ago). A summary of fossil insects is given by Kukalová-Peck (1991).

II. MAJOR DIVISIONS

Insects and their relatives (together referred to as the Hexapoda) are arthropods, with a chitinous exoskeleton and jointed appendages. They grow by molting,

Encyclopedia of Biodiversity, Volume 3

periodically shedding their exoskeleton to allow a new larger body to expand and harden. The closest arthropod relatives to the hexapods are still unknown but probably include some or all of the myriapods (centipedes and millipedes).

The insects formerly included all groups currently classified as Hexapoda, but recently authors have separated the Collembola (springtails) and Protura into a separate group called Ellipura or Parainsecta (Table I). The Diplura are placed either within the Insecta or in a separate order of uncertain affinity. All are soft-bodied hexapods usually found in soil or decaying organic material. Most are small, 1 to 5 mm in length, but some Diplura are much larger. The Collembola is the most species-rich group and the one most commonly encountered, as they are often enormously abundant in soil, compost, under rocks, in damp places, and even on the surface of water or snow.

The "true" insects are defined by characters of the structure and musculature of the antenna, among others. There are two relatively primitive groups, the jumping bristletails (Archaeognatha) and the bristletails or silverfish (Thysanura). Both are elongate, wingless creatures with a long, segmented, median "tail" and two shorter, but still prominent cerci. Both groups are scavengers, feeding on vegetable and sometimes animal debris.

The appearance of wings marks the Pterygota. The origin of wings and flight in insects has been controversial, with a number of theories put forward. Most now agree that wings are formed from a movable part of the of the leg, with its attendant corollary that the protowings were formed in aquatic insects, possibly used for ventilating gills and even for locomotion in water. Extant insects have wings on only the mesothorax and metathorax (the second and third thoracic segments), but fossils exist for which there are also prothoracic wings or wing precursors.

Also unlike the more primitive insects, pterygote insects have hemimetabolus development. The immature stages, often called nymphs, after molting several times, undergo a final molt that produces fully winged, sexually mature adults (except in the Ephemeroptera, which are unique in having a second adult molt; discussed later). More primitive hexapods are ametabolus, with gradual growth through several molts until the sexually mature adult stage, which differs relatively little from the immature forms, is reached.

The most primitive winged insects are sometimes joined in a group called Paleoptera and include the extant orders Ephemeroptera and Odonata. Unlike their more advanced relatives, the neopterans, paleopterans cannot fold their wings down against the body. Both extant orders of Paleoptera have aquatic larvae, but they differ markedly in other aspects of their life histories. Mayflies are generally herbivores, scraping algae and

TABLE I

Major Divisions of Extant Insects, Simplified

Superclass Hexapoda	
CLASS ELLIPURA (= PARAINSECTA)	
Order Collembola (springtails)	6,000
Order Protura	500
CLASS DIPLURA	
Order Diplura	800
CLASS INSECTA	
Primitive insect groups	
Order Archaeognatha (jumping bristletails)	350
Order Thysanura (bristletails, silverfish)	350
Pterygota (winged insects)	
Order Ephemeroptera (mayflies)	2,200
Order Odonata (dragonflies and damselflies)	5,000
Neoptera (advanced pterygotes)	
Order Plecoptera (stoneflies)	2,000
Order Blattodea (cockroaches)	4,000
Order Isoptera (termites)	2,300
Order Mantodea (mantids)	1,800
Order Grylloblattodea (ice crawlers)	25
Order Dermaptera (earwigs)	1,800
Order Orthoptera (grasshoppers, crickets, katydids)	20,000
Order Phasmatodea (walkingsticks)	2,500
Order Embioptera (webspinners)	200
Order Zoraptera	30
Order Psocoptera (bark and book lice)	3,000
Order Phthiraptera (lice)	3,000
Order Hemiptera (including Homoptera; true bugs)	98,000
Order Thysanoptera (thrips)	4,500
Endopterygota (= Holometabola)	
Order Coleoptera (beetles)	350,000
Order Megaloptera (alder flies, dobsonflies)	300
Order Raphidioptera (snakeflies)	175
Order Neuroptera (lacewings, antlions and others)	5,000
Order Mecoptera (scorpionflies)	500
Order Siphonaptera (fleas)	2,400
Order Strepsiptera (twisted-winged parasites)	500
Order Diptera (flies)	120,000
Order Lepidoptera (butterflies and moths)	160,000
Order Trichoptera (caddisflies)	7,000
Order Hymenoptera (sawflies, ants, bees and wasps)	120,000

[a] Important subdivisions are in bold. A detailed classification can be found in Kristensen (1991, Insects of Australia, Vol. 1). Approximate number of described species from various sources.

diatoms from objects in the water. When mature, they emerge from the water in an intermediate, winged stage called the subimago, which lasts only a short time. They molt once more and emerge as winged adults, ready to mate, and die soon thereafter. Dragonflies and damselflies, in contrast, have highly predaceous larvae. There is no subimago stage, as the adults emerge directly and begin relatively long lives as predators of smaller insects. Dragonflies in particular are strong fliers that can travel great distances from water when in search of prey.

The Neoptera are defined by the ability to swivel and fold the wings back at rest. The functional significance of such a modification is obvious—it allows the insect the opportunity to slip into small places to hide from predators and seek food. It has also opened the path to using the wings as protective structures when not in flight, and the forewings of many groups have become hardened or thickened. This is especially evident in the beetles, whose hardened forewings are largely useless for locomotion and are held out of the way during flight.

The relationships among the more primitive Neoptera are not well resolved. One particularly enigmatic group are the stoneflies (Plecoptera), whose aquatic lifestyle is reminiscent of the paleopterans. The larvae live in clear, cool streams or lakes, where they are herbivores or predators. The adults are weaker fliers than Odonata but longer lived than Ephemeroptera, and they feed on algae, decaying vegetation, or detritus.

The Blattodea, Isoptera, and Mantodea are terrestrial insects sometimes unified in a group called Dictyoptera. Their lifestyles are extremely divergent, however. Cockroaches are flattened, usually cryptic and nocturnal scavengers. They vary in shape and size from small, delicate species a few millimeters in length to large, bulky forms 6 cm long. Most are cryptically colored, but some are mimics of apparently distasteful fireflies (Coleoptera: Lampyridae).

Isoptera are termites, the only eusocial group of insects outside of the Hymenoptera. All species feed on cellulose from wood, leaves, and plant debris that is digested in their gut by a symbiotic microbial fauna of flagellate protozoans or spirochaete bacteria. Some species also culture gardens of fungi in their droppings; the fungi are then consumed and their enzymes aid in digestion. Termites have a number of castes, with most individuals in a colony being soft-bodied workers, along with a lesser number of soldiers and a single queen and king. Nests can be in the ground, in rotting wood or in carton nests built on trees or other structures. Ground-nesting species often build prominent mounds that in some species can reach a few meters in height. Although they appear defenseless, termites have a wide variety of chemical and structural modifications that allow them to resist attack by predators, especially ants.

The praying mantids, Order Mantodea, are voracious predators. They have modified, spiny, raptorial forelegs that grasp and crush their prey, usually other insects but sometimes even small vertebrates. They have no venom or other mechanism for killing their prey other than by eating them. Most mantids are green or brown in color, camouflaging them from attack by predators and detection by their prey. Some are flattened and specifically colored to resemble leaves or petals of flowers.

Unlike the well-known cockroaches, termites, and mantids of the Dictyoptera, the grylloblattids are poorly known and only relatively recently recognized as a separate order. They are found in cool climates, usually at high elevations, in the northern hemisphere living mostly under rocks or in caves. In the spring and summer they can forage on the snow surface, feeding on dead insects and plant material. Ice crawlers lack wings, which they have secondarily lost probably as an adaptation to their cryptic lifestyle.

Earwigs, order Dermaptera, are a distinctive group of insects that have a pair of forceps-like appendages at the posterior apex of the abdomen. Wings are absent in some species, and when present are modified, such that the forewings are shortened and hardened and the hingwings are extensively folded, exposing most of the abdomen. Most species are omnivorus, predatory, or herbivorus and are nocturnal. Usually they are encountered when searching through leaf litter, under rocks, in rotten logs, or in any other hidden crevice.

A much larger group is the Order Orthoptera, the familiar grasshoppers, crickets, and katydids. Most of these have notably enlarged hind legs, used to jump great distances or launch the insects in flight to escape predators. Many katydids are masters of camouflage, with green, leaf-like wings bearing markings that resemble leaf veins, fungal infections, and even insect feeding damage. Other orthopterans are equally cryptically colored, a testimony to the relentless selective pressure exerted by sharp-eyed bird predators. Some grasshoppers and katydids are strong fliers, whereas crickets are usually much more likely to stay on the ground. Most male Orthoptera produce sound to attract mates, and the calls of katydids and crickets are an integral part of the evening chorus in warmer parts (or seasons) of the world. Plants are the food of most species, although there are some predators as well.

Like the Orthoptera, the walkingsticks (Order Phasmatodea) are masters of concealment. With their long, stick-like bodies covered in warts, bumps, and spines, walkingsticks can virtually disappear in the appropriate background. One family, whose species are flattened but still extremely well camouflaged, is more appropriately referred to as the leaf insects. All feed on leaves of plants.

The webspinners of the Order Embioptera (webspinners) are an unusual group of small insects that live in silk galleries of their own production. They are narrow bodied, wingless (except for some males), and have the first tarsal segment swollen and filled with silk glands. The galleries they spin are under bark, beneath stones, or in the open in more humid regions; from them they feed on vegetable debris, rotten wood, moss, or lichens.

The Zoraptera are probably the least well known group of insects. They are nondescript, small insects, a few millimeters long, white or pigmented, usually wingless, and with or without eyes. Their life history is little known, although they may be fungus feeders.

A much more commonly seen, but still relatively obscure group are the Psocoptera, the bark and book lice. These small insects, under 1 cm in length, have large, globular heads, rounded bodies, and may or may not bear two pairs of wings. They live on plants, on bark, in leaf litter, or sometimes in human habitations; their food is plant, fungus, and dead insect debris.

Possibly related to some members of the Psocoptera, the Phthiraptera or lice are specialized ectoparasites of mammals and birds. Some species, such as the chewing lice, feed on skin, hair, or feathers, whereas others, such as the sucking lice, suck blood from their hosts. All are small, wingless, flattened insects found only on their hosts.

Outside of the Endopterygota (discussed later) there is only one truly large order of insects: the Hemiptera. This assemblage, in the broad sense, includes groups sometimes named the Heteroptera, or true bugs, and the Homoptera, the cicadas, leaf-, tree-, and planthoppers, aphids, whiteflies, scales, and others. All have distinctively modified mouthparts that are in the form of a piercing-sucking beak that they use to obtain food. Predatory species pierce their prey, usually other insects, injecting digestive enzymes to kill and begin the process of digestion. Predation is restricted to some heteropterans; a few also feed on vertebrate blood. Most Hemiptera are plant feeders, including the majority of the heteropterans and all the homopterans. Most species are terrestrial, although some heteropterans are aquatic. Some homopterans, especially scales and aphids, produce honeydew, as sweet excretion drips down onto leaves and becomes a major source of food for other insects such as flies and wasps. Some are attended by ants, who harvest honeydew and protect the homopterans from predators and parasitoids.

The thrips, Order Thysanoptera, are another unusual order of insects. They are small, slender-bodied, with or without slender, fringed wings. The last tarsal segment of the legs have an inflatable bladder used to improve the grip on the substrate. The mouthparts are asymmetrical, of the piercing-sucking form, used to feed on debris, fungi, or plants. Thysanopterans also have an unusual form of development in which there are two or three pupa-like stages, similar to the pupal stage of holometabolus insects, but which were independently evolved. At least one tropical species has developed a degree of sociality.

The major group of insects, accounting for about 80% of the species, is the Endopterygota, or Holometabola, two names that refer to different aspects of these insects. Endopterygota refers to the development of the wings from imaginal disks of tissue within the pupa of the immature insect, in contrast to the gradual, external development in less derived insects. Holometabola refers to the development of individual insects through a complete metamorphosis, through egg, larva, pupa, and adult stages. The larva is transformed to an often strikingly different adult through the pupal stage, in which the body tissues are almost completely broken down and then reorganized into adult structures. Imaginal disks, sections of tissue sequestered and formed in the larval stage, are retained in the pupa and are the basis of many of the adult structures, such as gonads, legs, and wings. The almost complete separation of the structure of larval and adult insects allows the holometabolus insects to be extremely specialized in each stage: the larva for feeding and the adult for mating and dispersal. This specialization increases the efficiency of each stage and probably results in the incredible success of holometabolus insects.

Within the Endopterygota, there are four orders that are immensely large and successful: the Coleoptera, Diptera, Lepidoptera, and Hymenoptera. Any one of these megadiversity groups alone is among the largest assemblages of organisms in existence, outnumbering any other group except plants. Together, these four insect orders constitute about 40% of all described species of life on earth.

The most successful order of insects in terms of species number is the Coleoptera or beetles. They are the largest group of organisms in the world, even outnumbering plants. The mesothoracic (first) pair of wings of the beetles are greatly strengthened and hardened, such that they are of little or no use in flight but are superb shields when held over the vulnerable abdomen. These hardened forewings, called elytra, are usually held flush over the back of the beetle, making it slippery and difficult to grasp, as well as hard and difficult to crush. In many groups, the overall body form is solid, flattened, and compact, allowing beetles to hide easily, penetrate cryptic habitats, and even to burrow extensively in soil. Although some beetles have secondarily evolved softened or even greatly reduced elytra, there is no doubt that this modification has allowed beetles to become the preeminent form of insect life on the planet. Primarily a group that is prevalent in warm climates, beetles become relatively less diverse at higher latitutes, but within their range they perform a number of roles, from scavenging, predation, and herbivory to endo- and ectoparasitism. Many species of beetles can be found under stones, in rotting logs, under bark, in freshwater, in fungi, in leaf litter, and on the foliage of plants. Some beetles found on flowers are mimics of bees and wasps, and some beetles found in army ant colonies resemble their hosts to an amazing degree. There are species of rove beetles found in termite nests that carry a passable model of a termite, formed by extensions of their abdomens, over their heads.

There are three orders of insects that are related to beetles but that do not share their same degree of success: the Megaloptera (alder flies, dobsonflies), Raphidioptera (snakeflies), and Neuroptera (lacewings, antlions, and others). The larvae of megalopterans are predatory and aquatic, usually living in clear, running water. The adults are medium to large-sized insects, some of which have enormously enlarged mandibles. In contrast, the larvae of snakeflies are terrestrial, living on the ground in rotting wood or leaf litter and feeding on smaller insects. Adults have a greatly elongated prothorax, giving them a snakelike appearance. Neuropterans are more diverse, with a number of elaborate or bizarre forms intermingled with the more normal-looking lacewings and the damselfly-like antlions and owlflies. All are relatively soft-bodied, with large wings bearing an elaborately reticulate wing venation. Some adults are known to be predatory, as are most of the immatures. The immatures of antlions dig conical pits into which prey fall and are grabbed by the waiting larva. Lacewing larvae are predatory on aphids and other soft-bodied insects found on foliage.

Another large assemblage of holometabolus orders includes the Mecoptera (scorpionflies), Siphonaptera (fleas), Strepsiptera (twisted-winged parasites), and Diptera (flies). The relationships among these orders is in a state of flux, with molecular data giving some indication that fleas are highly modified scorpionflies and that the Strepsiptera, long associated with the Coleoptera, are actually the closest relatives to the Diptera. Previously, some subdivision of the Mecoptera was thought to be most closely related to Diptera; whatever the situation, the monophyly of the scorpionflies relative to some other groups is seriously in question.

Scorpionflies are a relatively small group of insects, so-named because the male genitalia resemble the stinger of a scorpion. Most are terrestrial; the larvae are scavengers and the adults scavengers or predators on small insects. Adults and larvae of the unusual, brachypterous Boreidae feed on mosses, whereas the adults of one genus of Panorpodidae are herbivorous. One family, Nannochoristidae, has aquatic immatures that feed on larval chironomid midges. Some species have elaborate courtship behaviors involving nuptial gifts of dead insects to the female.

Fleas are highly modified, laterally flattened, wingless blood feeders on birds and mammals. Usually they are found in the nests or other areas frequented by their hosts. The larvae are usually free living, feeding on organic detritus and blood from droppings of the adults, although some are obligate ectoparasites. Adults are extremely laterally flattened, allowing them to travel smoothly between the hairs or feathers of their hosts, and often have backwards-pointing combs of spines that makes them difficult to remove. They have strongly developed jumping legs that allow them leap to and from their hosts.

Strepsiptera are extremely unusual parasitoids of other insects. The adults are highly sexually dimorphic, with the males being free-living, winged insects, whereas the females in all but one family are endoparasitoids—wingless, legless, and without only vestigial eyes and appendages of the head. The female body extrudes from the body wall of the host, emitting pheromones to attract males that copulate with the special openings (external genitalia being absent). The larvae are hypermetamorphic, with active, host-seeking first instars that, upon encountering and infecting a host, molt to a legless, relatively inactive form.

As one of the megadiversity groups described earlier, Diptera are found nearly everywhere. Their distinctive innovation has been the reduction of the metathoracic wings to a pair of knoblike halteres that act as gyroscopes in flight. This modification has increased their maneuverability and allowed Diptera to become unparalleled masters of aerial locomotion. Although some adult flies require extensive protein meals to mature eggs and power flight, most of the feeding is done by larvae, which can be predators, scavengers, herbivores, parasitoids and even true parasites. Free-living Diptera larvae are found in soil, rotting vegetation, feeding on and in plants, and sometimes exposed on vegetation. Aquatic forms are found in the silt or sand underlying the body of water (sometimes interstitially), on the surface of rocks, logs, or vegetation, or in the water column. Parasitoids attack mostly other arthropods, but some endoparasites attack mammals.

The orders Lepidoptera (butterflies and moths) and Trichoptera (caddisflies) are close relatives. The lesser of the two—the caddisflies—have aquatic larvae that are found in nearly every type of freshwater environment. Most construct cases or shelters from plant particles, twigs, stones, or sand grains tied together with silk. Some also construct nets to capture debris for food, whereas others are predatory, attacking other aquatic insects. The adults are slender, mothlike insects, often with long, thin antennae.

The Lepidoptera are among the best-known insects, especially the colorful, diurnal group called butterflies. Most of the diversity of the group, however, is in the nocturnal, often drably colored moths, which constitute about 80% of the species of Lepidoptera. The larvae are usually called caterpillars and are best known as voracious feeders on plants. Larval feeding takes place on the surface of the plant or within it (as in leaf miners and stem borers), and almost every plant part—leaves, stems, roots, flowers, and seeds—can be affected. Some species are also predatory and some feed on animal material, such as wool, but almost all species are phytophagus. Adults of most families have mouthparts that are modified to form a long, coiled tube that is used for taking up liquids, usually nectar from flowers. Many adult and larval Lepidoptera are cryptically colored to avoid detection by predators, whereas others are brightly colored to warn of toxic chemicals sequestered in their bodies. Some adults engage in long-distance migrations.

The final large group is the Hymenoptera, the sawflies, ants, bees, and wasps. Like the Diptera, the Hymenoptera have become extremely adept fliers, but have done so by joining the fore- and hindwings with a row of tiny hooks (hamuli) to produce a single functional pair of wings. The most primitive families are phytophagous, but from within these groups have arisen a great diversity of parasitoids, predators, and plant feeders. The parasitoids include some of the smallest known insects, which attack the eggs of their much larger hosts (such as Lepidoptera). Other parasitoids attack a diversity of immature insects, especially those of other holometabolus groups, and develop as endo- or ectoparasitoids. Some are obligatory hyperparasitoids—parasitoids of parasitoids; others oviposit in plant tissue and induce the formation of plant galls, in which the larvae feed. The predatory Hymenoptera attack a wide range of hosts, especially other arthropods, which they often subdue but do not kill with a venomous sting. Larvae of these species have a supply of fresh food to consume when they hatch from an egg laid on the paralyzed prey. Most species hide their prey in some sort of burrow or nest to prevent its being taken by other insects or scavenging animals. Some of these provisioning wasps have moved on to pollen and nectar are food, as in bees. Sociality has evolved a number of times in the Hymenoptera, with the largest and most complex colonies formed by ants and bees.

III. CONCLUSION

As this brief survey shows, the largest, most diverse and arguably most successful groups of insects are those that have undergone three major innovations: (a) the development of wings, (b) the ability to fold wings over the body, and (c) the division of the life history into four major stages. Further outstanding success was found in those species that had hardened forewings (beetles), those that transferred the responsibility of flight to a single pair of wings (Diptera), those that extensively exploited flowering plants (Lepidoptera), and those that combined a single functional wing complex with a number of other traits (such as haplodiploidy) that produced the ants, bees, and wasps (Hymenoptera).

See Also the Following Articles

BEETLES • BUTTERFLIES • FLIES, GNATS, AND MOSQUITOES • GRASSHOPPERS AND THEIR RELATIVES • HYMENOPTERA • INVERTEBRATES, TERRESTRIAL, OVERVIEW • ISOPTERA • MOTHS • TRUE BUGS AND THEIR RELATIVES

Bibliography

Boudreaux, H. B. (1979). *Arthropod Phylogeny, with Special Reference to Insects.* John Wiley & Sons, New York.

Hennig, W. (1981). *Insect Phylogeny.* John Wiley & Sons, Chichester.

Naumann, I. D. (Chief Ed.). (1991). *The Insects of Australia*, volumes I and II. Cornell University Press, Ithaca.

INTERTIDAL ECOSYSTEMS

A. J. Underwood and M. G. Chapman
Centre for Research on Ecological Impacts of Coastal Cities, University of Sydney

GLOSSARY

assemblage The collection of animals and plants found together in a patch of habitat.

diversity index A numerical measure combining information about the number of species present in a sample or habitat and information about their relative abundances.

ecological functions Attributes or properties of assemblages or habitats that are dependent on the biodiversity of animals and plants present. Examples are production of nitrogen, sequestering of heavy metals, sustained production of harvested resources, maintenance of a diverse food web, and so on.

ecological processes Direct or indirect interactions between species—such as grazing, predation, competition, responses to disturbance, and so on—that cause spatial or temporal patterns in distributions and abundances of species.

patchiness Variation from place to place (or time to time) in the abundances of animals or plants, caused by the interaction of numerous processes.

recruitment Arrival of new juvenile animals or plants into a habitat or an older stage of the population. For many marine animals, recruitment occurs some time after the settlement of planktonic larvae in the adults' habitat.

INTERTIDAL AREAS such as rocky shores, mangrove forests, sandy beaches, and mudflats are habitats for very diverse assemblages of plants and animals. These species are usually distributed patchily in space and time because of several important ecological interactions. Local biodiversity is therefore patchy and variable. Measurement of biodiversity is complicated by variability. The functions provided by biodiversity are not well understood in most intertidal habitats. Planning for conservation, management, and restoration of biodiversity is difficult because of variability in space and time.

I. INTRODUCTION

Intertidal habitats range from rocky beaches or platforms to sandy and muddy shores. They occur under a wide range of physical stresses, most notably due to emersion when the tide falls and disturbances due to waves. The animals and plants must endure the full range of physical variation, from being covered by water when the tide is in, to being potentially exposed to typical terrestrial conditions when the tide is out. In addition, the amplitude or range of tidal rise and fall can vary between two tides in a day (this is called "semidiurnal inequality") and, everywhere, varies in a fortnightly cycle between spring and neap tides. Spring tides rise much higher and fall much lower than do neap tides. As a consequence, during neap tides, animals and plants at high levels on a shore are not reached by an incoming tide for several days in a row and must endure long periods of inactivity or reduced activity until spring tides start again. At these high levels, moisture from spray (particularly where wave action is great) and during rain is very important for the supply of food and for preventing marine animals from drying out during their long periods of emersion.

In complete contrast, animals and plants at lower levels on the shore are submersed continuously for several days during neap tides because the tides do not fall low enough to expose such areas to aerial conditions. As a result, the animals and plants are, effectively, in a fully marine habitat and can be subject to predation by crabs, fish, starfish, and so on, which are mostly subtidal species. Such predators can forage at will over the lower parts of a shore during neap tides. This variation in periods of emersion versus submersion causes a very stark gradient of physical conditions from the top to the bottom of the shore.

A second strong physical gradient often exists along a shore, due to variation in wave action. As any shore bends away from being fully exposed to the open ocean, the strength of waves hitting the rocks will vary. As a result, there is a general decrease, with decreasing force of waves, in the amount of splash and spray over any area. The two physical forces—the direct force of waves and the ameliorating effects of splash and spray reducing the influences of desiccation during low tide—influence the abundances of animals and plants and the nature of their activities. Thus, grazing and predation are less intense where waves are strong. The diversity of species changes with wave action because of the combined influences of waves themselves, their ameliorating influence on desiccation, and the indirect influence of both physical features on the interactions among species.

These two major processes—tidal rise and fall and wave action—are so well known that their influences were long considered the major, if not the only, influences on local ecological processes and therefore diversity in intertidal habitats. The general features of intertidal regions, particularly the general patterns of distribution and the numbers of species on rocky shores, have for a long time been described in terms of these physical forces. Summaries and details are available in Lewis (1964) and Stephenson and Stephenson (1972) and other reviews of particular biogeographical regions or comparisons of various parts of the world.

More modern syntheses, in contrast, have focused on the ecological processes that have more profound effects in terms of creating and maintaining considerable patchiness in the ways animals and plants are distributed in intertidal habitats. The variation often is considerably more conspicuous than the features considered general and easily explained in terms of gradients of physical forces. This forms the central theme of this contribution—the processes that influence patterns of distribution and abundance of species and therefore the local diversity of species in intertidal habitats. The major focus is on rocky intertidal habitats, because most of the experimental work in the past 40 years has been on this type of habitat. The animals and plants are typically relatively short lived (although individuals can live for extraordinarily long periods). They are usually small and abundant, and most of them are slow moving or sessile. As a result, experimental manipulations to test specific hypotheses from competing models that might explain patterns can be done without enormous areas of habitat being disturbed and without excessive costs of logistics and infrastructure. There has been a mass of such work and it has revealed and continues to reveal a great deal about the life histories, interactions, and the general and specific ecology of the maintenance of diversity in intertidal habitats (see review by Underwood, 2000).

In addition to rocky shores, there is also some discussion of other intertidal habitats (sandy beaches, mangrove forests), even though these are discussed in detail elsewhere in this encyclopedia. Their inclusion here is to demonstrate that the sorts of variability in time and space found on hard, rocky surfaces are widespread and general and not confined to rocky shores.

Following an overview of ecological patchiness and some consideration of its causes in intertidal habitats, three consequences of variability in biodiversity are introduced: how to measure and compare diversity from one place to another, how descriptions of biodiversity may relate to functional aspects of diversity, and the sorts of functions that biodiversity may provide in coastal habitats.

II. PROCESSES CAUSING PATCHINESS ON ROCKY SHORES

Numerous ecological processes influence or maintain patterns of local biodiversity in coastal habitats. These include direct and indirect biological processes of recruitment, predation, grazing, and competition. They also include physical disturbances and biological responses to them. These different types of processes are illustrated here.

A. Recruitment

Variability in recruitment of coastal marine organisms is well described (Connell, 1985; Thorson, 1950). Many of the animals and plants have a relatively long dispersive phase in their life history, resulting in a large proportion of propagules being killed by predators, failing to develop or not being able to find a suitable habitat to settle and metamorphose. Inevitably, the processes of production and loss of propagules lead to great variation in numbers surviving to settle in the adults' habitat and the number actually settling in any area of habitat at any particular time.

There are two major consequences to issues of biodiversity of this variation in recruitment to adult populations. First, vagaries of types and numbers of species arriving in a patch will ensure that the diversity of species continues to be patchy at small spatial scales. Second, across the ranges over which species disperse, there will be little large-scale variation in diversity. Thus, along a stretch of coastline over which organisms disperse, the assemblages that develop will, in general, receive recruits from similar mixtures of species. At a local scale of patches of coastline in any area on the coast, there will be greater variation in the composition of assemblages.

At larger scales, speciation and changes in diversity from one side of an ocean to another are very much influenced by recruitment of larvae. Thus, the vast stretches of open, oceanic water in the middle of the Pacific and the Atlantic oceans form barriers to the spread or continuity of species. Scheltema (1971) sampled planktonic stages of life history of various gastropod mollusks in samples from the middle of the Atlantic. He found a very good correlation between the number of larvae of a given species and the distribution of adults. Species found on coastlines on both sides of the Atlantic were regularly found as larvae in his samples. Larvae that were found less commonly were from adults that had closely related or subspecies on the two coasts. This implied that lack of frequent interchange of larvae was causing differentiation and speciation. Species found on only one side of the Atlantic were rare or absent as larvae in the sample. Thus, where larvae cannot arrive, diversity will differ.

The timing of recruitment by different species can also affect local diversity of species. Thorson (1950) pioneered the study of influences of when recruits arrive to their final habitats. He described the predatory brittle stars (*Amphiura* spp.), which did not feed for 2 months while breeding. Several species of mollusks with short larval development bred during this period. The brittle stars had a longer period of development. As a consequence, the young mollusks could be released into the plankton, complete their development, recruit, and grow to a size where at least some of them would escape from the voracious predatory recruits of the brittle stars. If the mollusks recruited later, they were unable to avoid their predators. This anecdote needs to be examined experimentally but indicates that the timing and order of recruitment can have profound influences on which species survive in an area.

B. Competition

A second major process that can influence local diversity is competition—negative interactions among individuals of the same or different species caused by absolute or relative shortage of resources needed by all of the individuals. In most coastal habitats, resources are food, space, and, for plants, light. Space is needed to settle and become established (for sessile species) and as an indirect resource for food (for example, by mobile grazers, which need relatively open space over which to feed). Alternatively, sufficient space on the substratum is needed to allow access to the water column for filter feeders.

Generally, competitive interactions are considered to lead to reduction in abundances or actual elimination of competitively inferior species. Numerous examples demonstrate this, starting with Connell's (1961) classic experimental demonstration of destruction of one species of barnacle (*Chthamalus stellatus*) by another, faster-growing species (*Balanus balanoides*, using the taxonomy as in Connell). As a result of overgrowth by *B. balanoides*, *C. stellatus* could be eliminated from some lower areas of their vertical distribution on a shore.

Such competitive overgrowth can have dramatic effects on local richness of species. For example, Paine (1974) has described the major decrease in number of species living on primary space—the underlying substratum—where mussels are able to overgrow and even-

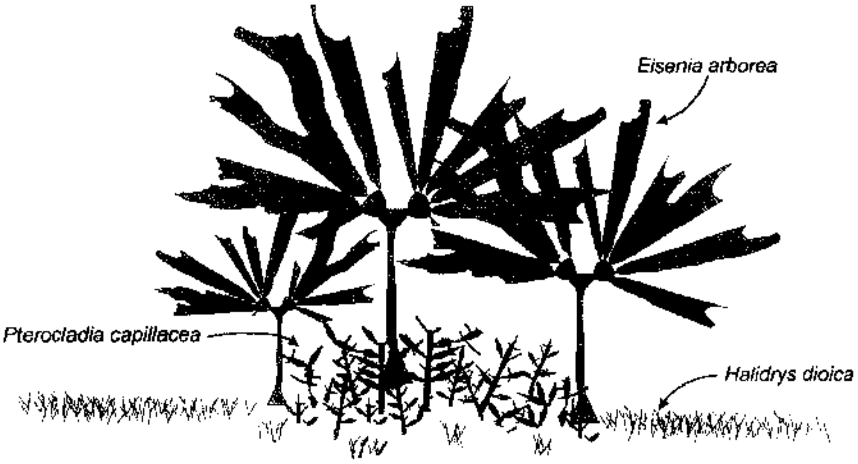


FIGURE 1 Diagram illustrating indirect interactions involved in competition for space and light among intertidal seaweeds (Kastendiek, 1982). The kelp, *Eisenia arborea*, forms a canopy that provides shade and moisture during low tide. Under the canopy, the red alga, *Pterocladia capillacea*, can thrive. Elsewhere, it is overgrown and outcompeted by the turfing alga, *Halidrys dioica*.

tually eliminate most other species (limpets, chitons, algae, barnacles).

These examples do, however, paint a biased picture. For example, where mussels overgrow a primary surface, their shells provide new hard substrata on which the inferior competitors can become established. They also serve as environmental engineers (Lawton, 1994), because the packed shells of mussels trap sediments. While feeding on particles in the water, mussels selectively ingest some sizes and types of particles. The rest are wrapped in mucus and eliminated as pseudo-feces, which add substantial substance and nutrients to the surrounding sediments in the matrix of mussels. The result of all these processes is provision of habitat and food for a very substantial number of species, such that the number of species in mussel beds is very large (Lohse, 1993).

There are also complications due to indirect interactions (reviewed by Wootton, 1994) where the outcome of competition between two (or more) species (say, A and B) influences the outcome of competitive interactions between one of them and yet other species (say, B and C).

For example, Kastendiek (1982) described the interesting case of the turfing red alga, *Halidrys dioica*, which outcompetes the alga, *Pterocladia capillacea*, by growing over and denying it access to light (Fig. 1). Where there is a canopy of a third species, *Eisenia arborea*, however, the inferior competitor survives well under the canopy and is able to "resist" competition from *H. dioica*. Thus, competition for light between *E. arborea* and *H. dioica* diminishes the latter's capacity to grow well and to take over space that is occupied by inferior competitors, such as *P. capillacea*. As a result, competition by *E. arborea* over *H. dioica* prevents competitive overgrowth of *P. capillacea* by *H. dioica*.

C. Predation

A major process influencing diversity of species on rocky shores is predation. Predation has three distinct influences. First is simply that where predators are sufficiently numerous or active, they may eliminate some species from areas of habitat. Thus, toward the bottom of rocky shores on the northwest coast of the United States, Connell (1970) described predation by several species of the whelk, *Nucella* (or *Thais*) feeding on barnacles. Low on the shore, the whelks had sufficient opportunity (due to prolonged submersion by the tides) to eat all of the barnacles, thus eliminating them from such low areas and thereby reducing the diversity of sessile species there. At higher levels, in contrast, the whelks were less active because they had a smaller period submersed during low tide in which they could find an item of prey and consume it.

The second influence of predation is indirect. The whelk *Morula marginalba* in southeast Australia consumes a variety of prey, including barnacles and small limpets, *Patelloida latistrigata*. It takes the whelks considerably longer to drill a hole through the shell of a barnacle than to eat a limpet; *M. marginalba* can consume *P. latistrigata* without having to drill a hole through their shells. Fairweather (1985) removed limpets in some areas as they arrived. As a result, predation of small barnacles by whelks increased. This demonstrated an indirect effect of preferences and differences in the time taken to consume prey; the least preferred species (the limpets) took less time to eat than the more highly preferred barnacles. Where there were no limpets, the whelks turned their entire attention to small barnacles, with consequent increased mortality and decreased abundances of the barnacles. Note also that whether or not limpets arrived from the plankton also influenced the eventual outcome in terms of abundances of barnacles and, therefore, the local relative abundances and, thus, diversity of species.

The final major influence of predation is also indirect. Large generalist predators are capable of consuming most other species in any patch of habitat. As a result, they can have direct influences on the abundances of prey and can free space on the shore for species to be able to recruit. Space is often in short supply, causing the various sessile species to compete with each other for space on which to live and causing grazing species to compete for space over which to feed. If the generalist predator disproportionately consumes "winners" of competitive struggles for space, predation will have very important effects on the local diversity of species. The original example of this phenomenon, known as "keystone predation," was investigated by Paine (1974). The large starfish, *Pisaster ochraceus*, eats most of the other sessile and mobile animals on rocky shores on the west coast of the United States. In some areas, the starfish preferentially eat mussels (*Mytilus* spp.). Mussels are, however, among the best competitors for space, being able to smother and grow over other species (see earlier). Predatory starfish continuously remove mussels, making space available for inferior competitive species that would otherwise be locally eliminated by overgrowth due to the mussels. In this case, the predators exert an indirect positive influence on a range of species of prey, resulting in there being greater numbers of species (i.e., greater species diversity) where there are predators than where there is none.

D. Physical Disturbances

Given that physical factors such as desiccation and wave action can have large and important influences on abundances and diversity of species in intertidal habitats, it is not surprising that physical disturbances can influence diversity. Where disturbances are large or frequent, it is likely that more delicate species that are unable to withstand the physical forces will be unable to survive. They are either prevented from settling and becoming established in the first place or are unable to complete their life cycles before a large disturbance kills them. It is expected under this model that there would be reduced diversity in more physically stressed habitats simply because of the direct loss of species due to disturbances. In contrast, where disturbances are small or rare, most, if not all, species can survive the rigors of the environment. Resources of space and food will then become critical and some species will disappear because of superior competitive abilities of other species. Under these circumstances, diversity in physically benign areas will also be reduced below what is theoretically possible.

As a consequence of these two processes, it is anticipated that the greatest diversity should be found in areas of intermediate disturbance, such that competition for resources is not too intense because there is sufficient disturbance to prevent all of the possible species from building up excessive abundances. At the same time, physical stress is not so great that some species are actually eliminated. This is the model of intermediate disturbance (Connell, 1978).

From this model, it has been predicted that in areas where physical disturbances can be increased or decreased in intensity or frequency, there will be predictable changes in diversity. For example, where disturbances are small, increasing their severity will cause increases in diversity as superior competitors are prevented from building up sufficient numbers to consume all of the resources needed by other species. If disturbances are increased even more, there will be a decrease in diversity.

Such predictions have been tested in several studies on rocky shores, with mixed success. Sousa (1980) identified patterns of diversity that were fairly consistent with the model, but the mechanisms operating were not as stated. Species were not lost by superior competition for space from other species. Rather, the life histories of the different species and the time taken to be able to recolonize disturbed areas were of greater importance. McGuinness (1987) only identified the intermediate disturbance model as being an appropriate explanation for observed patterns of diversity in a few of the cases be examined in intertidal boulder fields. Clearly, more work is needed to understand how and when the interactions of competition and disturbance lead to predictable outcomes in terms of diversity of species.

III. BOULDER FIELDS

Intertidal boulder fields form a very interesting habitat that forms part of rocky intertidal environments. Boulders have top surfaces that are exposed to waves and sunlight and that may be emersed during low tide. Underneath boulders, however, there is a very different sort of habitat—usually cool, moist, and dark. On shores affected by large waves, boulders are frequently overturned and moved. This alternately exposes each surface of the boulder to the light and causes the boulders to abrade against each other, damaging and killing many animals and plants living on them. These boulders support little biodiversity, mainly encrusting species that are resistant to harsh conditions or ephemeral species that rapidly colonize new habitat when it becomes available.

Boulders on sheltered shores, in contrast, are stable and support great diversity of animal and plant life. Different species are found on the tops of or underneath these boulders. Those on top are similar to those living on nearby rocky shores—after all, they are subjected to similar environmental conditions. Therefore, boulders may be covered with leafy and encrusting seaweeds, have patches of apparently bare rock, or support large numbers of snails, limpets, and other common intertidal animals.

The species living under boulders are completely different. There are few leafy seaweeds, but there may be many kinds of encrusting seaweeds, which seem to thrive in dark conditions. There can also be patches of attached animals—plate-like bryozoans, jelly-like masses of colonial ascidians or sea-squirts, encrusting sponges, and the calcareous and sandy tubes of different types of worms. These sorts of sessile animals are not common on rocky shores, except in rock pools or crevices or under ledges (Fig. 2).

Many of the mobile animals living under boulders are rather specific to boulders—they tend to be rare or absent from other intertidal habitats. Many show very specific patterns of behavior that allow them to move quickly from one boulder to another if the boulders are overturned or moved into unsuitable conditions by large waves. Therefore, although chitons on rocky shores tend to be very slow-moving and inactive animals, those under boulders move extremely rapidly over the rock surface or curl up and drop from the boulder as soon as it is overturned. This behavior returns them to the undersurface of the same or nearby boulders.

Intertidal boulder fields are potentially important habitats for the conservation of biodiversity. They often form rather small patches of habitat—especially when compared to sandy beaches, mangrove forests, or rocky shores, which are often continuous for many kilome-

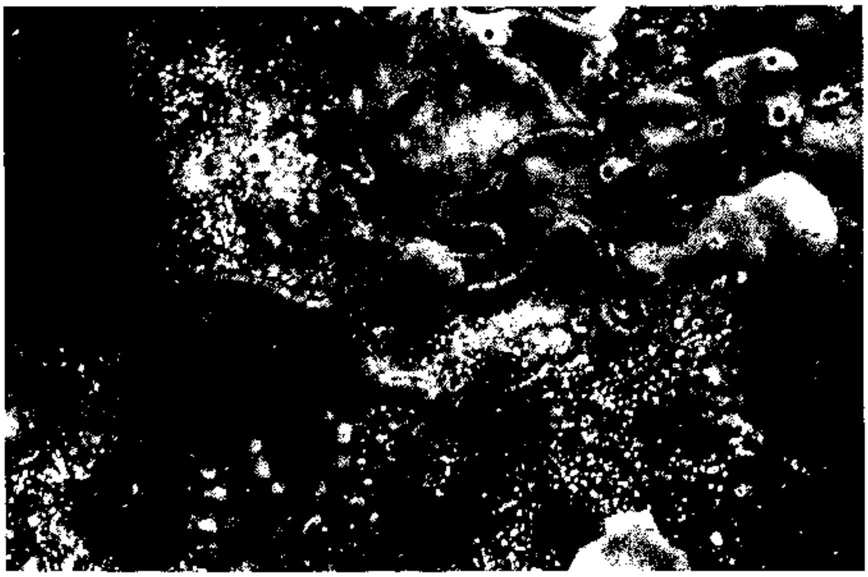


FIGURE 2 The undersurfaces of intertidal boulders support numerous species of sessile and mobile animals, many of which are not found in other intertidal habitats. These include encrusting sponges, ascidians, and bryozoans, numerous chitons and snails, and starfish, sea urchins, and brittle stars.

ters. They are also often scattered along the coast, with large stretches of other habitats between them. There are no data about how much interchange there is among populations in different boulder fields nor how consistently species would recruit from other boulder fields if locally eliminated. Because many species living in boulder fields are relatively confined to these habitats, they are likely to be vulnerable to human activities that destroy or damage their surroundings.

Finally, many species are extremely overdispersed. This means that each species tends to be crowded onto relatively few of the available boulders, while most boulders are unoccupied by that species. Because different species are often found on different boulders, patterns of biodiversity are very complex—both within and among different boulder fields.

As yet, the processes that lead to these complex patterns are not well understood but include such factors as the size and shape of the boulder, what stone it is composed of, where it is (e.g., depth of water, what substratum is beneath it), and the different wave conditions where the boulder field is located. In addition, there are complex interactions among the animals and plants themselves (Sousa, 1980), causing complex changes to biodiversity depending on which species happen, by chance, to recruit to which boulders.

IV. MANGROVE FORESTS

In contrast to intertidal rocky shores and boulder fields, mangrove forests are dominated by trees. Unlike some other habitats containing many large plants—for example, subtidal kelp beds (Foster and Schiel, 1985) and terrestrial rain forests (Terborgh, 1992)—patterns of diversity in mangrove forests (i.e., variation in the numbers and types of animals and plants found from place to place or time to time) are not well documented (Hutchings and Saenger, 1987). This is particularly true for any measures of small-scale patchiness in diversity within individual forests. This contrasts markedly with other intertidal habitats in which patchiness of animals and seaweeds is relatively well described (see the discussion on rocky shores presented earlier).

This lack of good quantitative measurements of biodiversity in mangrove forests is due to many different factors. First, the plants that make up the forests are, themselves, not very diverse. Typically, they do not develop a complex structure of many species of canopy, understory, and ground cover, which could, in turn, support many different types of animals. Second, because mangrove forests normally grow on very sheltered shores, the vertical gradients of environmental conditions associated with tidal height are very strong (Hutchings and Saenger, 1987). There are also strong latitudinal gradients, causing very large differences between temperate and tropical zones in the types and diversity of trees that make up the forests. These are well documented elsewhere and are not discussed further here. Within any single mangrove forest, however, different species of trees typically live at different levels along the intertidal gradient, so that in any one place, even fewer types of plants are found.

The main factor that has led to poor description of the distribution and patchiness of biodiversity in mangrove forests is, however, that most diversity is due to small invertebrate animals. This is, of course, true for many habitats where much of the biodiversity is "invisible." In such habitats, appropriate sampling can clearly measure variation in diversity from place to place or time to time, but these sampling designs are inevitably complex and costly because they need to measure diversity at many different spatial and temporal scales.

The animals that make up most of the biodiversity in mangrove forests include many species of crustaceans, such as crabs and amphipods, small snails and bivalve mollusks, worms from many different phyla, insect larvae, and so on. Because these animals are generally very small and live on or under the surface of the mud, they are not readily visible without sampling small patches of mud, sieving the animals out of the sediment, and viewing them under magnification. The arduous work that this entails means that patterns in diversity are not well described for most mangrove forests.

In addition, because the types of plants found in mangrove forests often vary along the intertidal gradient, most studies of the distribution and diversity of animals have also described similar broad-scale patterns. Therefore, changes in the diversity of snails, crabs and other animals from the seaward to landward edges of the forest tend to dominate the literature. There have been few descriptions of the small-scale variability or patchiness in this diversity within a shore level, although this may be the most important source of variation in diversity, as is the case on rocky shores. Some studies that have been done to measure small-scale patchiness in diversity in mangrove forests show that much of the variability in diversity and abundances of these animals is at very small spatial scales (i.e., among patches of habitat centimeters or meters apart). It can also change unpredictably through time (Fig. 3).

This sort of patchiness of diversity is not limited to little invertebrates that live in the mud. Many of the larger animals that spend their juvenile stages in man-

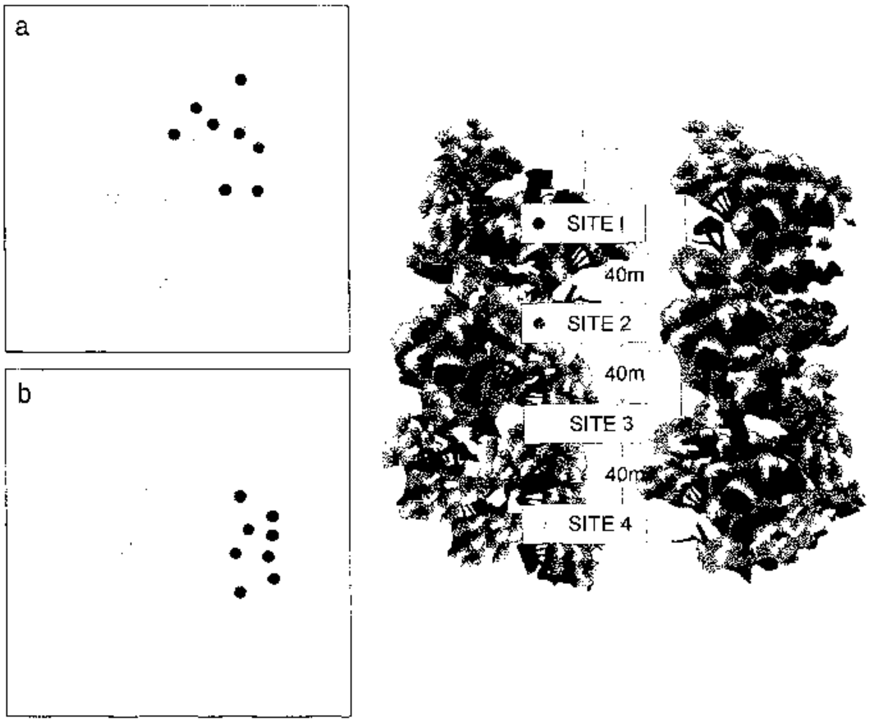


FIGURE 3 This figure shows variability in the numbers and types of animals living in a mangrove forest in New South Wales, Australia. The animals were sampled in four sites along the bank of a river. Each site was sampled using four quadrats, placed 1 to 2 meters apart. Each point on this figure represents the types and numbers of animals in each quadrat. The relative distances between two points indicates how similar were animals in those two quadrats. Points close together represent quadrats containing very similar types of animals. Points far apart mean that the quadrats contained very different numbers and types of animals. Figure 3a shows that when the animals were sampled the first time, the diversity of animals was very variable from one quadrat to another. Quadrats in the same site (meters apart) showed similar amounts of variability in these measures of diversity as shown by quadrats in different sites. Figure 3b shows what the patterns looked like only three months later. Animals were more similar at the scale of meters (among quadrats in each site), but the sites appeared to represent a gradient along the river. This relatively large change in the pattern of diversity was not seasonal, predictable, nor found in other nearby sites. It was mainly due to changes in the relative numbers of numerous small crustaceans.

grove forests, such as fish, prawns, and large crabs, show similar patterns. A recent survey of small fishes in patches of mangrove forests in Sydney Harbour showed most of the variation in abundance and diversity was found at the scale of meters (i.e., from one net to another set in a small patch of mangrove forest). Variation from site to site (hundreds of meters apart in the same bay) and from bay to bay (kilometers apart) together accounted for less variability.

In contrast to rocky shores, there is little information about processes that cause this small-scale patchiness in mangrove forests. This is due to the fact that, with few exceptions, mangrove forests have not been as well studied, particularly using well-designed, controlled field experiments. Typically, diversity of small animals living in sediments in any aquatic habitat is primarily determined by the range of grain sizes of the sediment itself. Larger species tend to be more common in coarse-grained sediments and vice versa. The range of sediment in any particular mangrove forest can vary from patch to patch because of such processes as changes to water currents around submerged objects (e.g., the bases of trees) or biological processes.

In New South Wales (Australia), the mangrove crab, *Heloecius cordiformis* creates mounds of coarser and drier sediments and flat areas of wetter and finer sediments when feeding and burrowing (Warren and Underwood, 1986). Although not yet measured, it is likely

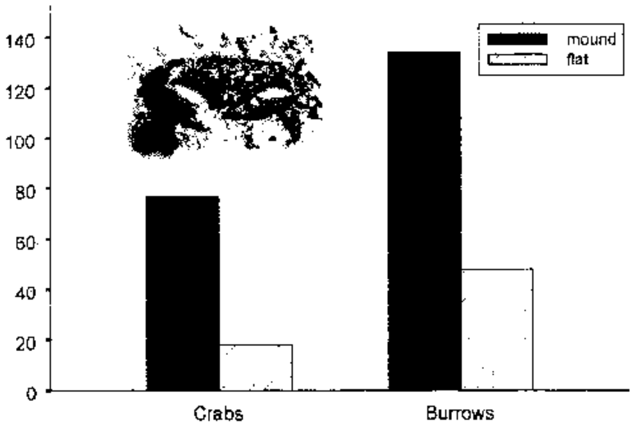


FIGURE 4 The crab, *Heloecius cordiformis*, and their burrows are more numerous on mounds of sediments than in the flatter patches of habitats in between the mounds. These mounds are created by the crabs themselves From J. H. Warren, unpublished data.

that this activity directly influences local patterns of diversity, as shown for the burrowing behavior of soldier crabs in Tasmania (Fig. 4) (Warwick *et al.*, 1990). Other small-scale structures—buried shells, the roots of the mangrove plants, worm tubes—are equally likely to alter patterns of diversity, although the importance of these small structures and the behavior of other animals as influences on biodiversity are not yet well documented for mangrove forests.

V. SANDY BEACHES AND MUDFLATS

Sandy beaches and mudflats tend to grade into each other along a continuum depending on the relative grain size of the sediment. Sandy beaches have rather coarse-grained sediments and mudflats are, of course, muddy, with fine-grained sediments. Both can occur on very sheltered shores, such as in estuaries, but mudflats do not occur on wave-exposed shores, where sandy beaches may be common (Brown and McLachlan, 1990; Reise, 1985).

Sandy beaches and mudflats have very few large plants, although heaps of decaying plant material, commonly known as beach wrack, can be found on the strandline. This material—decaying seagrasses, seaweeds, and kelps that have been washed up by the waves—is rapidly decomposed by bacteria, fungi, and small animals, many of which are dependent on wrack for food. This activity ultimately releases nutrients back into the coastal waters (Griffiths *et al.*, 1983).

Sandy beaches and mudflats appear at first glance to resemble deserts—there are often few animals to be seen, other than birds and a few crabs. The birds are generally visitors to these shores, feeding along the shoreline or over the extensive flats during low tide, but spending the rest of their time elsewhere. There are relatively few animals high on the shore, mainly semiterrestrial species, such as ghost crabs which spend the day in burrows above the level of high tide, coming out at night to feed. Nevertheless, despite first appearances, there is diverse life lower on the shore. As is the case in mangrove forests, however, most of the animals live under the sediments.

Really small animals live in the tiny spaces among the grains themselves. These include unicellular and multicellular animals from many different phyla. Many of these very small animals resemble juveniles of larger species. Many are also extremely simple. For example, although multicellular, they may only consist of a few cells and lack structures—such as limbs, kidneys, or a circulatory system—relying on exchange of substances through the body wall to acquire oxygen or get rid of waste (Brown and McLachlan, 1990).

The larger animals living under the sand or mud either construct and maintain complex tubes or burrows, or simply move through the sediment without leaving any structure behind them. On sandy beaches exposed to relatively large waves, the sediment is continually reworked by wave action. Most animals living in these habitats do not construct tubes or burrows because these are likely to be destroyed. Instead, they remain relatively deep in the sediment or tend to have mechanisms whereby they can rapidly burrow back under the sediment if exposed. Mole crabs have paddle-shaped limbs and can burrow extremely quickly during the time available between waves. Similarly, many bivalves have an extendible, mobile foot, which rapidly digs and anchors into the sediment and then contracts, pulling the bivalve under the sediment in only a few seconds.

Because mudflats tend to be very extensive and not steeply sloping, they usually remain quite damp during low tide. Many different species of worms, crustaceans, mollusks, and so on can be found in these habitats. Although some simply burrow through the sediment, feeding as they burrow on detritus and particles attached to the grains, others live in semipermanent tubes or burrows. These can be very complex. Some burrows of crabs have many entrances and exits and are connected together in a complex underground labyrinth. Some animals living in burrows emerge to feed on the surface of the mud itself during high tide or at night. Other species remain in the burrow, but they may have long tentacles that are spread over the surface of the mud to pick up particles of food. Animals inhabiting burrows that feed on particles in the water often have modified limbs or other appendages, which can create currents of water through the tubes, pulling in the food with the water current. They also have very elaborate filters, which sieve the particles of food out of the water. These filters are modified legs, mouthparts, gills, or other organs.

Despite the fact that sandy beaches and mudflats are very common intertidal habitats, their patterns of biodiversity are not particularly well known. Available data suggest that, like most other intertidal habitats, they are extremely patchy and variable through time. The numbers and types of animals differ from one patch to another only a few meters away. These patterns are caused by the different species responding to subtle changes in the sediments themselves, to local disturbances such as feeding by fish which creates depressions in the mud and to other animals living in the sediment via interspecific processes such as competition or predation.

Some species have rather aggressive behavior that spreads them out, presumably to ensure adequate food for all. Other species tend to aggregate and can create large patches of tubes or burrows, altering the sediments around them. These sorts of processes inevitably change small-scale local patterns of diversity, but there is very little information on how such processes may operate. Of course, as described earlier, such processes modify the initial patterns of diversity, which are established by differential recruitment of species among different patches of habitat.

VI. MEASUREMENT OF BIODIVERSITY IN COASTAL HABITATS

Methods for and interpretations of measures of biodiversity in coastal habitats have proven complex. As reviewed by Gray (2000), the simplest measure—the number of species in any area—does not describe the structure and variation in assemblages. For this, measures are needed that can collate information on the types of species present (the composition of an assemblage) with information about their relative abundances. Many indices have been used (see Magurran, 1988). One of the most widely used is the Shannon-Wiener index:

$$H' = -\sum_{i=1}^{s} p_i \log_2 p_i$$

where p_i is the proportion of species i in the whole sample. Following Whittaker (1972), Magurran (1988) described measurements of diversity in a single sample as "point diversity"—that in a set of replicate samples from the same habitat as "alpha diversity." Comparisons from habitat to habitat, for example, along an environmental gradient uses measures of "beta (or between habitat)" diversity. This is the combined diversity across a set of alpha diversities. There are also larger-scale measures, such as "gamma diversity," which are measured across larger spatial areas, such as a whole coastline.

There are serious problems of scale in these considerations. For example, a rocky headland could be a relatively large area, so diversity measured across it would be gamma diversity. Within the area, there are different habitats such as rock pools, algal beds, boulder fields, and so on. In each of these, samples would give measures of alpha diversity. Beta diversity would measure the diversity in, say, algal beds along a gradient of increasing exposure to waves.

Suppose, instead, that the study was a series of headlands along the biogeographical gradient from south to north along a coastline. It is now likely that the diversity measured in a set of randomly taken replicates from one headland would be considered to be alpha diversity. That from the set of headlands would be beta diversity. So whether or not a sample allows measurement of alpha or gamma diversity is entirely dependent on the spatial scale being examined! It is clearly crucial to have defined very carefully the hypotheses being tested so that only the relevant scales of habitats and sample areas are examined.

Another major difficulty in measures of diversity is that the number of species found is almost always dependent on the size of the sample examined or the length of time spent searching for individuals throughout the habitat. This is inevitable because of two distinct aspects of the patchy distributions of species. First, rare species are not going to occur very often in any sample and will often be missing from small samples. A lot of space needs to be covered to be sure that the rare species are actually encountered. Imagine a comparison of the animals in two habitats, which both have one hundred species in them. If in one habitat (A) many more of the species are rare than in the other (B) and similarly sized small samples are examined from each habitat, there will be apparently fewer species in habitat A than in B. This is, however, an artifact of sampling.

A second reason for there being inevitable differences in the numbers of species seen in two samples where one is much larger than the other is due to potential heterogeneity of habitats in the area studied. Suppose that the number of species is recorded in samples totaling 4 square meters of habitat in one area and 8 square meters in the other habitat. The latter sample would normally be expected to contain more species if there are many subhabitats in the area sampled. So on rocky shores, clumps of seaweeds, small rock pools, patches of mussels, and so on may be scattered at scales of tens of centimeters. Each of these subhabitats probably contains several different species, in addition to "cosmopolitan" species found throughout the entire area. As a result, the larger area examined in samples from one habitat will encompass more different patches and therefore contain more species than would be found if a smaller area were examined.

There exist techniques to compensate for differences in the size of sample from one habitat to another. One of the most commonly used is called rarefaction, but this is known to have serious problems of overestimation of numbers of species in small samples (Fager, 1972). A superior alternative is to take random smaller samples from the largest, to match the size of the smaller samples from the other habitat. Such random samples can be used to give reliable estimates and estimates of error associated with the number of species expected in smaller samples. Comparisons between the two habitats then consist of comparisons of the entire sample from the habitat that was sampled with smaller samples and the randomly determined estimates from the other habitat (see the review by Gray, 2000).

Some multivariate methods are widely used to compare the diversity of species from one area or habitat to another. These operate on the principle that the abundances or biomasses of every species can be recorded from a series of replicate sample units (cores, grabs, quadrats, etc.). This produces a matrix of species by replicates for a sample from each habitat or area. The difference from any one sample unit to another can then be calculated across all of the species. One popular measure of such difference is the Bray-Curtis dissimilarity index (Clarke, 1993). This is calculated as follows:

$$D_{jk} = \frac{\sum_{i=1}^{s} |X_{ij} - X_{ik}|}{\left(\sum_{i=1}^{s} X_{ij} + \sum_{i=1}^{s} X_{ik}\right)} \times 100$$

where D_{jk} is the dissimilarity between sample units j and k, X_{ij} is the abundance of species i in sample j, X_{ik} is the abundance of species i in sample k, and s is the total number of species found in the two sample units. If j and k come from the same sample, this is a measure of variation among units within a sample. If j and k come from two different samples, D is a measure of dissimilarity between the two samples. Because there are replicate measures, there is a sample of measures to estimate variation within each of the two samples and between the two samples. Statistical tests can then be done to determine the likelihood of there being more difference between two samples than expected by chance given the variability within the two samples (Clarke, 1993).

Such procedures have been used widely in examinations of the effects of environmental impacts and other ecological studies on diversity of animals in benthic habitats, where diversity is made up of the composition of species in any habitat and their relative abundances. They are limited in the complexity of sampling and experimental designs for which they can be used, so that natural variability at a variety of spatial scales and from time to time are difficult to examine using these procedures. There are, however, developments using much more intensive sampling to obtain large enough samples that independent measures of dissimilarity can be used in complex designs to test hypotheses about spatial and temporal scales of variability in entire diverse assemblages (Underwood and Chapman, 1998).

VII. CAN INTERTIDAL BIODIVERSITY INDICATE ECOLOGICAL FUNCTION?

Although it is clear that "biodiversity" covers many levels of variability, from genetic to population diversity within species, from diversity among species to diversity among habitats or ecosystems, the most common perception of biodiversity is still simply the numbers and types of animals and plants in different areas. Therefore, the importance of biodiversity has often focused on terrestrial habitats, the consensus of opinion being that there are more species on land than in the sea. Nevertheless, at different levels of organization, marine systems are more diverse—for example, there are more phyla in the sea than found on land (Gray, 1993). In addition, marine species appear to have more genetic diversity than related terrestrial and freshwater organisms. This suggests that they may be less vulnerable to processes that reduce genetic variation, although there is increasing concern that many nearshore aquacultural practices, including culture of fish and bivalve mollusks, are decreasing genetic variability of natural populations.

Although many small marine invertebrates are still to be identified and described, it is generally held that most diversity in the sea is of benthic organisms. There are many more species of animals and plants living on or in close association with the seafloor, rocky reefs, and others, than live in the water column itself. This is probably related to the diversity of different benthic habitats. There is still controversy about the relative importance of deep-sea habitats and coastal habitats to benthic diversity—controversy arising from the fact that these habitats are so poorly described that general patterns of biodiversity across large areas of marine habitat, even nearshore and intertidal habitats, are largely guesswork.

At whichever level biodiversity is examined, however, understanding patterns of biodiversity is essential to understanding the role(s) of diversity in the maintenance of different ecological functions. By functions, we mean such processes as maintaining diversity itself, recycling nutrients and chemicals among organisms and their surrounding habitat and maintaining appropriate habitat for adequate food, shelter, and so on. Because patterns of biodiversity are complex—biodiversity varies at many spatial scales from centimeters to hundreds of kilometers and changes predictably and unpredictably among days, weeks, seasons, years, and eons—the role of diversity in maintaining the functioning of individuals, populations, assemblages of species, and habitats will also be necessarily complex.

Relationships between biodiversity and the persistence of successful ecological functioning may be described by quite different theories. First, there may be considerable redundancy of species, so that many species may be lost without impairment of function because other species simply take on their roles. There obviously must be a limit to this—not all species can disappear and ecology persist. A second view is that all species may contribute to any ecological function, but the system may withstand loss of species until a crucial limit is reached, after which there will be rapid and inevitable degradation of function. The third view, called the idiosyncratic response, is most likely. This states that there will be loss of function with loss of species, but this is likely to be variable and unpredictable because of the complex patterns of abundance and interactions among the various component species.

Most of the experimental tests of relationships between biodiversity and ecological function have been done in terrestrial systems. The results of the different experiments are not clear-cut and there is considerable controversy about their designs and the ways in which they have been interpreted (Huston, 1997). Therefore, even in those systems that have been examined experimentally, there is no unambiguous relationship between biodiversity and any ecological functions. Because of fundamental differences in the ways that different processes operate between marine and terrestrial habitats, it is extremely unlikely that the results of such experiments, even if not controversial, could be applied to marine systems. Therefore, it is essential that the importance of biodiversity to the well-being and persistence of coastal habitats be tested experimentally in those habitats. Without such tests, it is not possible to predict what might be the results of any changes to diversity of these flora and fauna.

To date, the functional role of marine biodiversity has largely focused on off-shore processes. For example, people are concerned about diversity of plankton and its role in large-scale processes, such as commercial fisheries or atmospheric levels of CO_2 and O_2. In coastal habitats, most emphasis has been on the role of diversity in maintaining ecological function and has concentrated on single species or suites of similar species. Therefore, there is concern about the loss of large filter-feeding bivalves in estuaries and bays because the large numbers of these animals were considered to filter all of the water in the estuary in only a few days. In areas where populations have crashed, the water has become more turbid and is now occupied by different types of animals than was the case in the past. Similarly, certain large predators or grazers on intertidal shores or shallow reefs appear to have a pivotal role in maintaining patterns of biodiversity, by selectively removing competitively dominant sessile animals or plants with their associated species, thereby proving space for a host of other organisms that thrive in the altered conditions (see earlier discussion on keystone predators).

Ecological engineers (Lawton, 1994) have also received considerable attention because of their roles in altering the physical environment, creating habitat and changing the availability of resources to other species. Many coastal animals are ecological engineers—the most obvious of which are corals which may develop into large reefs—a unique and very diverse habitat totally dependent on the corals themselves. Other lesser-known examples include burrowing bivalves and crabs in salt marshes, which can alter drainage, sedimentation, and erosion in addition to having direct impacts on other animals or plants which they may eat or for which they may be an important source of food. Some intertidal animals form habitat for a range of other species. Therefore, mussels of the genus, *Mytilus*, provide habitats for more than 300 other species of animals and plants on intertidal shores in Washington in the United States (Suchanek, 1992).

Mangrove trees stabilize sediments, modify shorelines, prevent erosion, and provide habitat for the juveniles of commercially exploited fish and shellfish. Many studies examining the relationship between structure and function in mangrove forests have concentrated on changes in the number of trees, especially in areas where they are cut down for firewood.

There has, however, been little consideration about the role of the real biodiversity of such habitats—the mud-dwelling bacteria and small animals—on the well-being of the trees themselves or, indeed, on the maintenance of the habitat itself. Unfortunately, as described earlier, little is known of this biota, other than the fact that it is variable, unpredictable, and complex. Because the species are poorly described and even more poorly understood, it is generally accepted that there is a lot of redundancy in the system. Therefore it has been assumed that, if the larger plants are looked after, the small invertebrate animals and the functioning of the habitat will look after itself.

The same is true of all intertidal habitats. Most ecological functions are thought of in terms of general, widespread, and large-scale processes (e.g., the cycling of nutrients throughout a mangrove forest or the maintenance of food webs across a rocky shore). Yet the fauna and flora that make up most of the diversity and that maintain these ecological functions are variable and patchy and change unpredictably through time, often at very small and localized scales. The scales at which ecological structure (the species) and ecological functions (many of the processes) are measured are often quite different. It is therefore difficult to relate the stochastic variability that we see in ecological structure with what is often perceived as the predictability in measures of ecological function. It is increasingly important that we measure structure and function at a similar range of scales before we will be able to understand the role of natural patchiness in biodiversity in ecology. This is not yet widely done, but it is crucial if we are to conserve functioning habitats in an increasingly altered world.

VIII. WHAT SERVICES DOES INTERTIDAL BIODIVERSITY PROVIDE?

Consideration of relationships between diversity of species and functioning of ecological systems leads rapidly to consideration of any "values" of biodiversity. Because biodiversity is not well described (i.e., the threats to it are poorly understood and probably variable from place to place and habitat to habitat), there is still considerable controversy about perceived value(s) of biodiversity. This is specifically true for most coastal habitats, especially those on temperate coasts, which have not received as much attention from the media and celebrities as have habitats, such as coral reefs or tropical rain forests.

There are many so-called values to biodiversity, some of which are discussed in Box 1. Whatever terms are used, however, they tend to fall into two main catego-

Box 1

Despite the fact that there is no unambiguously accepted list of what is meant by the values of biodiversity, many values are cited as being important.

Evolutionary value. This includes genetic diversity that may allow organisms to persist or change in response to localized threats to diversity. This may be very important in intertidal and nearshore coastal habitats where humans tend to cause most damage to marine habitats.

Ecological value. A naturally functioning habitat with its full complement of biodiversity is necessary for studying and understanding ecological patterns and processes. This understanding is, in turn, considered essential for the persistence of life on earth in the face of humanity's alteration to and degradation of the natural world.

Economic value. There are or may be species of direct economic value. On intertidal shores, this includes plants and animals collected or cultured in fisheries, mangrove trees exploited for firewood, and so on. Habitats that are used for recreation may also have economic value. For example, a wetland with diverse species of birds, can have economic value if bird-watchers are prepared to pay to access it.

Aesthetic value. Natural habitats are sometimes perceived to have value simply because they exist. This is also sometimes called "existence value."

Ethical value. There is argument by many sections of society, that all biodiversity has ethical value—that is, we, as human beings, are ethically bound to protect biodiversity. This argument comes to the fore when decisions are being made about the rights of humanity to eliminate diseases and other pests from the earth.

ries. Ecocentric values are those associated with the well-being of the animals and plants themselves (i.e., they are focused on maintaining the full range of functioning ecological processes). They include such terms as ecological value and evolutionary value (see Box 1). Anthropocentric values, in contrast, are centered around humanity and the role of diversity in maintaining or improving human lifestyles. These obviously include a range of values, based on our perceived moral obligations, our ideas of the sort of world that we want to live in, and obvious economic returns from our exploitation of coastal resources.

In contrast to many terrestrial habitats, nearshore coastal habitats, including nearly all intertidal areas, are generally conserved for anthropocentric values rather than ecocentric values. Following traditions lasting hundreds of years, the sea is still largely regarded as a larder to be plundered rather than a unique set of habitats to be conserved. Apart from certain coastal wetlands, which can support large populations of wading birds, most intertidal habitats do not abound with large charismatic megafauna. Therefore, although they may contain threatened species, these are unlikely to attract media attention.

Although large areas of mangroves can be extensive, diverse, and very interesting forests, they are still generally valued for their economic returns—provision of firewood, habitats for juvenile commercially exploited fishes, sites for aquaculture, and, unfortunately in many countries, prime real estate for reclamation. Similarly, rocky reefs, mudflats, estuaries, and so on are generally considered important for what direct economic value they can provide humanity.

The loss or degradation of nearshore habitats goes largely unseen because the biota that live in them are small, cryptic, and unknown. Yet they are incredibly diverse, and in many parts of the world they make up a large amount of the endemic fauna and flora. Most of these organisms do not have commercial value. As long as these coastal habitats continue to be considered valuable mainly with respect to the direct economic services that they provide to humankind, there will be little initiative to do the research and impose the necessary management to conserve this diversity.

IX. CONCLUSIONS

The management of biodiversity in intertidal coastal habitats needs a different approach from that used in terrestrial habitats. Except for species that are considered to have a key role in local ecological processes, it is unlikely to be useful to attempt conservation on a species by species level. Exceptions are the so-called keystone species or ecological engineers, which were discussed in some detail earlier. Either these species provide habitat directly or their activities serve to enhance local diversity because of their elimination of competitively dominant species. For the most, however, emphasis on conservation of individual species is unlikely to be profitable.

First, for most species, we know too little about their requirements for habitat, food, and so on to be able to impose sensible managerial options. Second, we know too little about their natural patterns of variability in time or space to be able to evaluate whether any managed populations are persisting with anything like their natural patterns of abundance. Third, we certainly know too little about their interactions with other species to know which are essential to their continued well-being and which are not (i.e., which suite of species to try to conserve). In many cases, it has so far proven impossible to determine the geographical range of habitat occupied by a breeding population of intertidal animals, the extent to which there is interchange from one population to another or the sources of recruits to any area that must be conserved.

Management of coastal biodiversity will be better focused on habitat. Removing or reducing disturbances to patches of mangrove forest, rocky shores, mudflats, and beaches is one of the more feasible and cheaper options—it is usually easier to prevent access to areas than try to control people's behavior once they are in them. The fauna and flora will then be left to do the best they can under the circumstances. Fortunately, many species living in shallow-water, marine habitats are able to deal with many disturbances to their habitat, as long as these are neither too frequent nor too extreme. Evidence suggests that many can relatively rapidly recolonize areas once any disturbances are removed; such areas can then develop what appear to be functional ecological systems. Whether all species return to such habitats is not known because it is generally not known what species were there before the disturbance.

Nevertheless, there is obviously need to do more research on interactions among species wherever possible, so that keystone or engineering species can be identified. Where possible, management that will conserve these directly is likely to conserve those species that are dependent on them. Therefore, legislation that protects rocky shores from foragers removing mussels and other large shellfish will not only protect the targeted species but also a diverse range of other animals and plants. There is little chance that these could be protected on a species-by-species basis because too little is known about any individual species to mount a case for its protection.

Finally, with increasing understanding of the landscape ecology of coastal habitats and how "intertidal landscapes" interact, conservation of intertidal biodiversity is probably best within a mosaic of patches of different habitat. There is considerable debate about the value of few, large versus many, small reserves for conservation of terrestrial species. Which may be best is unknown for marine habitats and, in any case, there is no reason why what suits one species or set of species will suit another. The best strategy will probably be to hedge one's bets and try a range of different procedures. Therefore, with the current lack of understanding, conserving patches of habitat of different sizes and shapes, set at different distances apart and subjected to as wide a range of environmental conditions as possible, will probably be the safest option until we know a lot more about our intertidal biodiversity.

Clearly, our understanding of the processes that influence and change biodiversity is increasing at a satisfactory pace. Our ability to use these insights in planning and management of conservation is, however, more limited. The natural variability due to numerous processes and the large ranges over which larvae seem to disperse combine to make predictions difficult. The urgent tasks for coastal systems in response to threats due to coastal development, aquaculture, disposal of wastes, and potential rises in sea level require more understanding of scales of variability in distributions and abundances of coastal species.

See Also the Following Articles

COASTAL BEACH ECOSYSTEMS • MANGROVE ECOSYSTEMS • MARINE ECOSYSTEMS • MEASUREMENT AND ANALYSIS OF BIODIVERSITY

Bibliography

Brown, A. C., and McLachlan, A. (1990). Ecology of Sandy Shores. Elsevier, Amsterdam.

Clarke, K. R. (1993). Non-parametric multivariate analyses of changes in community structure. *Aust. J. Ecol.* 18, 117–143.

Connell, J. H. (1961). The influence of interspecific competition and other factors on the distribution of the barnacle *Chthamalus stellatus*. *Ecology* 42, 710–723.

Connell, J. H. (1970). A predator-prey system in the marine intertidal region. I. *Balanus glandula* and several predatory species of *Thais*. *Ecol. Monogr.* 40, 49–78.

Connell, J. H. (1978). Diversity in tropical rain forests and coral reefs. *Science* 199, 1302–1310.

Connell, J. H. (1985). The consequences of variation in initial settlement versus post-settlement mortality in rocky intertidal communities. *J. Exp. Mar. Biol. Ecol.* 93, 11–46.

Fager, E. W. (1972). Diversity: a sampling study. *Am. Nat.* 106, 293–310.

Fairweather, P. G. (1985). Differential predation on alternative prey and the survival of rocky intertidal organisms in New South Wales. *J. Exp. Mar. Biol. Ecol.* 109, 135–156.

Foster, M. S., and Schiel, D. R. (1985). The ecology of giant kelp forests in California: A community profile. *U. S. Fish Wildl. Serv. Biol. Rep.* 85, 1–152.

Gray, J. S. (1993). Marine biodiversity: Patterns, threats and conservation needs. *Biodivers. Conserv.* 6, 153–175.

Gray, J. S. (2000). The measurement of marine species diversity, with an application to the benthic fauna of the Norwegian continental shelf. *J. Exp. Mar. Biol. Ecol.* 250, in press.

Griffiths, C. L., Stenton-Dozey, J. M. E., and Koop, K. (1983). Kelp wrack and the flow of energy through a sandy beach ecosystem. In *Sandy Beaches as Ecosystems* (A. McLachlan and T. Erasmus, Eds.), pp. 547–556. W. Junk, The Hague.

Huston, M. A. (1997). Hidden treatments in ecological experiments: reevaluating the ecosystem function of biodiversity. *Oecologia (Berl.)* 110, 449–460.

Hutchings, P. A., and Saenger, P. (1987). Ecology of Mangroves. University of Queensland Press, St. Lucia, Queensland.

Kastendiek, J. (1982). Competitor-mediated coexistence: Interactions among three species of benthic macroalgae. *J. Exp. Mar. Biol. Ecol.* 62, 201–210.

Lawton, J. H. (1994). What do species do in ecosystems? *Oikos* 71, 367–374.

Lewis, J. R. (1964). *The Ecology of Rocky Shores*. The English Universities Press, London.

Lohse, D. P. (1993). The importance of secondary substratum in a rocky intertidal community. *J. Exp. Mar. Biol. Ecol.* 166, 1–17.

Magurran, A. (1988). *Ecological Diversity and Its Measurement*. Croom Helm, London.

McGuinness, K. A. (1987). Disturbance and organisms on boulders. II. Causes of patterns in diversity and abundance. *Oecologia (Berl.)* 71, 420–430.

Paine, R. T. (1974). Intertidal community structure: Experimental studies on the relationship between a dominant competitor and its principal predator. *Oecologia (Berl.)* 15, 93–120.

Reise, K. (1985). *Tidal Flat Ecology*. Springer-Verlag, Berlin.

Scheltema, R. S. (1971). Larval dispersal as a means of genetic exchange between geographically separated populations of shallow-water benthic marine gastropods. *Biol. Bull.* 140, 284–322.

Sousa, W. P. (1980). The responses of a community to disturbance: The importance of successional age and species life histories. *Oecologia (Berl.)* 45, 72–81.

Stephenson, T. A., and Stephenson, A. (1972). *Life Between Tidemarks on Rocky Shores*. W. H. Freeman and Co., San Francisco.

Suchanek, T. H. (1992). Extreme biodiversity in the marine environment: Mussel bed communities of *Mytilus californianus*. *NorthWest Environ.* J. 8, 150–152.

Terborgh, J. (1992). *Diversity and the Tropical Rain Forest*. Scientific American Library: distributed by W. H. Freeman, New York.

Thorson, G. (1950). Reproductive and larval ecology of marine bottom invertebrates. *Biol. Rev.* 25, 1–45.

Underwood, A. J. (2000). Experimental ecology of rocky intertidal habitats: What are we learning? *J. Exp. Mar. Biol. Ecol.* 250, in press.

Underwood, A. J., and Chapman, M. G. (1998). A method for analysing spatial scales of variation in composition of assemblages. *Oecologia (Berl.)* 117, 570–578.

Warren, J. H., and Underwood, A. J. (1986). Effects of burrowing crabs on the topography of mangrove swamps in New South Wales. *J. Exp. Mar. Biol. Ecol.* **102**, 223–235.

Warwick, R. M., Clarke, K. R., and Gee, J. M. (1990). The effect of disturbance by soldier crabs, *Mictyris platycheles* H. Milne Edwards, on meiobenthic community structure. *J. Exp. Mar. Biol. Ecol.* **135**, 19–33.

Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon* **21**, 213–251.

Wootton, J. T. (1994). The nature and consequences of indirect effects in ecological communities. *Ann. Rev. Ecol. Syst.* **25**, 443–466.

INTRODUCED PLANTS, NEGATIVE EFFECTS OF

William G. Lee
Landcare Research, New Zealand

GLOSSARY

economic impact Capacity of weeds to limit productive use of terrestrial environments and the costs associated with their control in both managed and natural areas.

environmental impact Ability of weeds to displace native fauna and flora, alter key ecosystem functions, and change disturbance regimes.

introduced plants Species transported by humans to regions outside their natural geographic range.

invasive weed Introduced plant species that establish self-maintaining populations and spread, with and without human assistance, into new areas where they frustrate human intentions in production and natural landscapes.

PLANT SPECIES THAT ARE either deliberately or inadvertently transferred by humans to habitats outside their native geographic range and which subsequently naturalize and have significant negative economic consequences and environmental impacts.

I. INTRODUCTION

Human migration, settlement, and trade over the past two centuries have produced a rapid globalization of floras with independent evolutionary histories that were previously kept apart by natural barriers. The scale of movement and mixing of floras in recent years is unprecedented in the earth's history, and very few parts of the world currently retain an exclusively native assemblage of plant species. Hundreds of naturalized plants, no longer dependent on humans for their persistence, occur in most regions and countries, where they form a significant and growing proportion of the wild flora. Over large continental areas the contribution of established introduced plants can reach 23% of the total flora (e.g., Canada), but this is greatly exceeded on many islands where the naturalized flora may outnumber the pool of native species (e.g., Bermuda, Ascension, New Zealand). Despite the many benefits provided by introduced plant species, there is growing concern at the environmental damage and economic costs of naturalized invasive plants in both productive and natural ecosystems. Important negative impacts are created by a small subset, probably less than 0.01% of all introduced species growing in a region, but their control, management, and impact consume significant economic resources, especially of developed countries.

The extent, patterns, and consequences of introduced plant invasions are described in another chapter presented in this volume, "Plant Invasions." This chapter highlights the negative impacts of introduced plant species using case studies of notable invasions, particularly those that have spread into natural areas with serious deleterious consequences for native species and communities. Examples of invasive plant species have been selected on the basis of a well-documented, and therefore relatively recent, spread and impact in natural environments. They are not necessarily the most widespread plant species, or the most aggressive, and some are limited to a few regions or countries. However, they are representative of the pattern and scale of intercontinental transfer of introduced plant species and the types of dramatic impacts these species can have when released into new environments. The examples chosen include a range of plant growth forms, altitudinal and bioclimatic zones, regions of origin, and modes of negative impact.

For ten introduced invasive plant species, representing algae *(Undaria pinnatifida)*, ferns *(Salvinia molesta)*, grasses *(Bromus tectorum, Spartina alterniflora)*, small herbs *(Hieracium pilosella)*, large herbs *(Fallopia japonica)*, cacti *(Opuntia)*, shrubs and small trees *(Lantana camara, Mimosa pigra)*, and tall trees *(Pinus pinaster)*, an account is given of the invasion process; local ecological, cultural, and economic impacts; attempts at control; and likely long-term consequences. Not all invasions are irreversible, and several examples illustrate successful levels of population control, if not eradication.

II. NOTABLE EXAMPLES

A. Undaria pinnatifida (Laminariales)

Some of the most pronounced ecological impacts and intractable problems of introduced species invasion involve marine organisms. Both deliberately and inadvertently introduced macroalgae have become invasive in recent decades with the increase in marine farming, transoceanic and coastal shipping, and the discharge of ballast water in and around foreign ports. The Asian laminarian kelp *(Undaria pinnatifida)*, native to Japan, Korea, and China, has been introduced and spread since 1970 in other northern hemisphere waters (e.g., England, French Atlantic, and in parts of the Mediterranean near France, Spain, and Italy). More recently (1980s), populations have been discovered in the southern hemisphere (Argentina, Australia, and New Zealand), making it the first large kelp to cross tropical waters successfully. In its native range *Undaria pinnatifida* is widely cultivated as an edible seaweed, and new locations are often viewed positively for their commercial potential. However, the alga is highly invasive, has significant negative conservation impacts on the local native shallow marine fauna and flora, and may be impossible to eradicate.

Native to the north-west Pacific, *Undaria pinnatifida* has an annual, heteromorphic life cycle, with macroscopic sporophytic and microscopic gametophytic stages. The golden brown, pinnatifid sporophyte, up to 3 m long, grows over autumn and winter, before degenerating in the summer. When growth slows, sporophylls develop as undulating extensions at the base of the stipe and release millions of short-lived (generally less than 2 days) male and female pelagic spores which, following settlement, develop into small gametophytes that may overwinter before producing either eggs or sperm. Fertilization of the female egg cell by male gametes gives rise to the straplike juvenile sporophyte. *Undaria pinnatifida* grows from low water to a depth of 20 m, depending on turbidity and wave exposure, and tolerates sea surface temperatures from 0 to 27°C. In Japan, it is an important edible seaweed crop and is also cultivated extensively to feed juvenile abalone (*Haliotis spp.*).

Undaria pinnatifida first appeared outside its native range in 1971 in a lagoon on the south coast of France, in the Mediterranean Sea. The source is thought to have been contaminated oyster spat translocated from the Pacific. It has subsequently spread along northern parts of the Mediterranean from Spain to Italy. In 1983 *Undaria pinnatifida* was deliberately introduced to the French Atlantic for experimental aquaculture trials, on the basis that sea temperatures would be too cool (< 14°C) for the release of zoospores. This assumption was proven incorrect when ecological surveys during the 1990s revealed natural establishment at five of the nine cultivation sites, some now abandoned, with several new populations up to 20 km away from nearest putative sources. *Undaria pinnatifida* appeared on the south coast of England in 1994, the population mostly likely deriving from microscopic gametophytes or juvenile sporophytes attached to boats arriving from Brittany.

Spread of *Undaria pinnatifida* into the southern hemisphere has been very recent, beginning in the 1980s, and accidental, presumably from plants attached to ships or in seawater ballast. *Undaria pinnatifida* was first discovered in New Zealand in 1987, and has subsequently spread, mainly via sporophytes and gametophytes attached to hulls of boats, to more than ten localities, up to 1000 km apart (Fig. 1). Soon after (1988) the kelp was discovered in Tasmania, in and around several major ports, and was recently (1996) reported from coastal waters on the Australian mainland. The first South American population was discovered in Argentina in 1992.

While fouling boats and underwater structures is of nuisance value, the full ecological impact of *Undaria pinnatifida* in these new habitats is hard to predict, as

FIGURE 1 Intertidal kelp forests of *Undaria pinnatifida* recently established along parts of the eastern South Island, New Zealand. Photo: National Institute of Water and Atmospheric Research. See also color insert, Volume 1.

the species is a very recent invader in foreign waters, is mostly confined to highly modified habitats in harbors, and initially preferentially colonizes clean, artificial surfaces. Investigations in the Atlantic suggest that *Undaria pinnatifida* is preempted on denuded rocky substrates by several local native annual and perennial kelps. Because of this competition, it may eventually be restricted along the coast in northern France to the low tide zone. In Tasmania, following disturbance and dieback of the native kelp due to ocean warming, *Undaria pinnatifida* displaces *Macrocystis* kelp forests and dominates subtidal urchin barrens due to its high fecundity and fast growth rate. However, it is less successful against native seaweeds on exposed coasts with slow currents. In New Zealand, *Undaria pinnatifida* reduces habitat for native marine grazers by smothering coralline algal surfaces favored for larval settlement of the animals. The region has a depauperate flora of annual and early successional seaweeds, and *Undaria pinnatifida* has few effective competitors on fresh surfaces created by storms and echinoid grazing. Frequent wave induced substrate disturbances could consolidate the dominance of *Undaria pinnatifida* in many shallow marine habitats.

While the transoceanic and local spread of *Undaria pinnatifida* has been dependent on human activity, the alga has several attributes that make it an aggressive competitor in shallow marine habitats. Sporophytes densely colonize any unvegetated surfaces, and huge populations can establish within a year on boats, wharf piles, mooring buoys, and on breakwater or reclamation debris. These can provide multiple sources for short- and long-distance dispersal. Importantly, *Undaria pinnatifida* occurring outside its native range can, in warmer waters, produce several generations of sporophytes each year, while local kelp species are reproductively dormant for part of the year.

The negative effects of *Undaria pinnatifida* on marine aquiculture and numerous algal communities are potentially enormous, and most countries where the species has established are pursing some form of control. However, the large reproductive potential and cryptic dispersal phases of Asian kelp, together with the partial obscurity provided by the marine environment, combine to make detection and control of *Undaria pinnatifida* almost impossible. Reducing the rate of spread of *Undaria pinnatifida* may be practicable with improved ship hygiene practices in ports and the strict control of aquaculture ventures. Eradication of small populations, using direct removal of sporophytes by divers, or sterilization of mariculture equipment, may also be possible where pioneer plants are restricted to artificial structures (e.g., mooring lines), but this is unlikely to be achieved where *Undaria pinnatifida* has established in harbors and along exposed coasts.

B. Salvinia molesta (Salviniaceae)

Ferns are rarely successful weeds in natural ecosystems, but a notable exception is *Salvinia molesta* (kariba weed, African pyle, Australian azolla, water fern, giant azolla), a free-floating perennial aquatic fern native to tropical and subtropical areas of South America. In the later part of the 20th century it has been spread worldwide, becoming invasive in fresh waters in warm-temperate to tropical areas, and causing major disruptions to the utilization of important water resources.

Only formally described in 1972, *Salvina molesta* is a sterile pentaploid, probably of natural hybrid origin. Native to parts of Brazil and Argentina, it grows there in lagoons, swamps, and river margins with still or slow-moving freshwater, in a diverse community of floating and emergent plant species. In these situations *Salvinia molesta* occurs at low densities, does not form extensive mats, and is rarely invasive.

Morphologically the species is quite plastic, exhibiting several distinctive forms. At each node *Salvinia molesta* supports submerged (which function as roots) and floating leaves. The size, shape, and density of the foliage changes depending on environmental conditions (temperature and nutrient concentrations) and growth phase. When *Salvinia molesta* first establishes, usually in calm, sheltered waters, the plants are small (20 mm), with oval, flat floating leaves. With time, leaves become rounded and larger, and in the final mat-forming stage are folded, crowded, and between 50 and 60 mm across. *Salvinia molesta* is efficient at accumulating and reutilizing nutrients such as nitrogen.

Salvinia molesta has become widely distributed through aquarium and horticultural industries and is now naturalized in Australia (Fig. 2), Africa, Madagascar, India, Sri Lanka, Southeast Asia, Philippines, Indonesia, Papua New Guinea, Fiji, and New Zealand. Dispersal between water systems is also invariably by human activities (e.g., boats, aquaria), but spread within and between interconnected water bodies is assisted by wind and water movement of vegetative fragments. Outside its native environment, *Salvinia molesta* shows phenomenal growth rates, via the expansion of auxiliary buds at stem nodes. Under optimum conditions (30°C and high nitrogen levels), plant area can double every 8 days and biomass every 2.2 days. In Australia total floating biomass values of 1600 g^{-1} m^{-2} dry weight, or 400 t ha^{-1} fresh weight, have been recorded. Frequently *Salvinia molesta* forms mats 1 m deep across extensive waterways.

During the 1970s and 1980s, *Salvinia molesta* was seen as potentially one of the worst invasive weeds globally, with major ecological and socioeconomic impacts. The basis for this concern is well illustrated by accounts of the species invasion in Papua New Guinea. In the early 1970s *Salvinia molesta* became established in the Sepik River flood plain, and within a decade covered 250 km^2 of water surface. The weed impeded water transport, preventing fishing, trade between villages, and access to medical and educational facilities. Fresh water for humans, stock, and wildlife became restricted, and *Salvinia molesta* mats proved suitable habitats for insect vectors of serious human diseases such as malaria. Similarly, in Zimbabwe, *Salvinia* threatened utilization of Lake Kariba, a massive impoundment on the Zambezi River created in the 1950s for hydroelectricity generation, commercial fishing, recreation, and transport. Within 15 years of the scheme being completed, mats of *Salvinia molesta* were limiting human access, decreasing fish populations, and modifying nutrient regimes. Elsewhere, *Salvinia molesta* has also invaded rice fields, blocked irrigation ditches, eutrophied small water bodies, and smothered emergent, floating, and submerged indigenous aquatic plants. A severe reduction in light and dissolved oxygen, together with an increase in carbon dioxide and hydrogen sulphide, have had a drastic impact on most benthic biota.

FIGURE 2 Attempted mechanical removal of extensive mats of *Salvinia molesta* blocking large waterways in Queensland, Australia. Photo: Queensland Department of Natural Resources. See also color insert, Volume 1.

The management and eradication of *Salvinia molesta* focused initially on mechanical removal and the use of chemical herbicides, and these were occasionally effective, especially where original populations were small or isolated. However, repeated treatment was required to prevent the rapid recovery of the plants, and the techniques used were expensive and occasionally environmentally damaging. Since 1980, biological control of *Salvinia molesta*, using a foliage- and stem-feeding weevil (*Cyrtobagus salviniae*—Curculionidae) from its native habitat, has been outstandingly successful in suppressing populations and achieving control in many countries (e.g., parts of Australia, Papua New Guinea, Fiji, South Africa). For example, the weevil had reduced the cover of *Salvinia molesta* on two 16 ha dams in Zimbabwe from over 90% to under 2% in less than two years. Similarly impressive results were achieved at a much larger scale in the Sepik River. Following the weevils' introduction, large mats of *Salvinia molesta* are typically reduced to sparse plants around the margins of water ways, interspersed with other aquatic plant species of the region. After the collapse of the *Salvinia* plants, weevil numbers also decline but appear to remain at densities sufficient to keep the weed in check. The global benefit of this successful example of biological control has been estimated to surpass $A200 million.

C. Bromus tectorum (Poaceae)

The development of rangelands for pastoral use, through stock grazing, fire, and the deliberate and incidental introduction of new plant pasture species, has involved some of the most extensive modifications to natural ecosystems ever undertaken by humans. In the semi-arid intermontane west of North America more than 400,000 km^2 of native sagebush-steppe communities have been penetrated or displaced this century by

Bromus tectorum (cheatgrass, downy brome, downy chess, drooping brome), a winter annual grass native to arid Europe and Central Asia. The dominance of this species has resulted in changes in land use, increases in erosion potential, reductions in native biodiversity, and alterations to the major disturbance regimes over large areas. *Bromus tectorum* is considered to be the most significant plant invasion in North America.

The native habitat of *Bromus tectorum* is centered on dry continental climates. Commonly associated with human activities, it occurs in grazed and ungrazed grasslands, among crops, and along roadsides. Flowers are hermaphroditic and self-fertile, and populations survive the dry summer period as seed. It characteristically germinates in autumn, estivates during the coldest parts of winter, and can grow rapidly, to around 50 cm tall in spring, when it flowers. The plants die back during the hot, largely rainless, summer. Seeds can be widely dispersed via animals, as they have long awns that are frequently embedded in fur or wool. Transfer by humans is usually accidental as a seed contaminant and among farm machinery. *Bromus tectorum* is now found in temperate East Asia, North and South America, and Australasia.

Prior to European settlement, the natural vegetation of the Great Basin area contained perennial cespitose grass-dominated steppes, the grasses (*Agropyron, Festuca, Stipa, Poa*) and forbs on mesic sites, giving way to shrubs (*Artemisia, Chrysothamnus*) in drier areas. The accidental introduction of *Bromus tectorum* into the intermontane region of the Pacific Northwest late in the 19th century, and several deliberate attempts to establish it as an alternative pasture species, coincided with large-scale domestic grazing by cattle, sheep, and horses. Because of the low abundance of large ungulates (bison and deer) during the Holocene, the original grasses and shrubs were generally poorly adapted to grazing ruminants. The flora also lacked aggressive colonizing annuals that could reclaim sites following disturbance. Burning by early settlers to remove shrubs and improve stock access further depleted the vegetation. Initially a weed of cultivated fields and roadsides, *Bromus tectorum* rapidly spread via wind and animal dispersal, and in seed lots, into overgrazed and vulnerable ungrazed native communities, becoming a widespread and dominant weed in the 1930s and reaching its current distribution in North America by the 1950s. Spread since then continues but at a slower rate.

The expansion of *Bromus tectorum* has caused multiple changes in natural ecosystems, largely through its impact on the grass/fire cycle. Accumulation of its highly flammable dry litter following spring growth, has resulted in an increase in the frequency of natural lightning-induced summer fires from 2 to more than 20 per century. Frequent burning has killed native shrubs and depleted the recruitment of perennial grasses. In contrast, *Bromus tectorum* benefits from fire, due to its large seed bank, prodigious and prolonged germination, rapid seedling growth, and heavy annual flowering. It is therefore able to preempt the recolonization of burned areas by native species already weakened by regular fires. Greater intensity and frequency of burning following *Bromus* invasion reduce plant cover, giving rise to erosion-prone landscapes. *Bromus tectorum* provides poor-quality herbage for stock, and its dominance has resulted in a decline in rangeland quality (Fig. 3).

FIGURE 3 Extensive arid steppe in Washington State, USA, dominated by *Bromus tectorum* which, following fire and grazing, displaces native shrubs (foreground) and grasses. Photo: Richard Mack.

The outcome for conservation and pastoralism of the transformation, in less than half a century, of natural semiarid steppe dominated by shrubs and perennial grasslands, to communities predominantly of introduced annual grasses, remains uncertain. *Bromus tectorum* appears uncontrollable under extensive management systems. It is still spreading into more isolated areas of the Great Basin and into the Great Plains region of the United States.

D. Hieracium pilosella (Asteraceae)

Numerous flat-weeds (rosette species) have accompanied European colonization of temperate regions around the world. Most of these plants are small, short-lived (2 years), dependent on human disturbance for local persistence, and generally inconsequential as weeds outside intensively managed systems. *Hieracium* *pilosella* (mouse-ear hawkweed), a perennial, stoloniferous, mat-forming daisy native to Europe, has a worldwide distribution in temperate regions and was generally considered to be relatively benign until it started to invade grasslands in New Zealand in the 1950s. Currently it occupies about 1 million ha of montane-subalpine short tussock (*Festuca novae-zelandiae*) grassland in subhumid and humid zones and is seen as the greatest threat to the conservation and long-term pastoral use of native grasslands. Key ecological factors responsible for the spread of *Hieracium pilosella*, and management options for its control, are controversial, and highlight the problems of large-scale weed management in nonforest ecosystems.

Native to Europe, *Hieracium pilosella* is a prostrate rosette-forming herb that reproduces vegetatively via stolons, which are initiated following flowering. Seed can be produced by both sexual and apomictic processes. In addition, hybridization and polyploidy are common and give rise to complicated population structures. Rosettes produce an erect short scape, which supports a solitary lemon-yellow flower that produces abundant small wind-dispersed achenes. In its native range the species is characteristic of disturbed dry habitats, most often grazed pastures, wastelands, and outcrops, particularly on calcareous soils below 1000 m above sea level.

Introduced into New Zealand in the middle of the 19th century, most likely as a seed contaminant, *Hieracium pilosella* remained locally restricted and largely insignificant for more than a century. The invasive phase of *Hieracium pilosella* became obvious during the 1960s, initially in the semiarid intermontane basins in the South Island, where the species became dominant in grazed *Festuca-Poa* grassland induced by pastoralism. Following burning of the original native forest and woodland during early Polynesian settlement, approximately 600 to 1000 years ago, tall tussock *Chionochloa* grassland and shrub land extended onto these areas. Early European pastoralists, in the first few decades of the 19th century, increased fire frequency and brought in sheep and cattle which, assisted by periodic eruptions of introduced rabbit populations, destroyed dominance by the tall cespitose grass species. The decline in plant cover and increasing bare ground provided numerous sites for invasive grazing-tolerant plant species. Foremost amongst these has been *Hieracium pilosella*, which slowly occupied the most depleted areas (those with 500 to 1200 mm rainfall) over about 50 years, becoming widespread and dominant in the 1970s when it began to extend into subalpine tall tussock grassland. In the 1990s *Hieracium pilosella* became the most widespread and significant introduced weed in native grasslands (Fig. 4), accompanied in parts of its range by the introduced congeners *Hieracium praealtum* and *Hieracium lepidulum*.

During its colonizing phase *Hieracium pilosella* forms distinctive circular patches that gradually coalesce to cover large areas. In eastern parts of the South Island, instantaneous rates of increase of approximately 8% per year have been recorded, with local cover values of *Hieracium pilosella* exceeding 50%. Much of this local expansion is attributable to vegetative reproduction, with seedlings being rare. *Hieracium pilosella* smothers and displaces small native herbs and grasses and restricts the survival of productive pasture grasses and legumes. Soil changes beneath *Hieracium pilosella*, especially increasing rates of acidification, also limit the growth of introduced legumes.

There is little doubt that human activities, especially repeated fire, and the grazing of domestic and feral animals, have increased the vulnerability of indigenous grasslands to invasion by introduced plant species. However, some relatively unmodified native habitats, especially in subalpine areas, on braided river beds, in semiarid environments, and on unstable hillslopes, may also be vulnerable due to tectonically induced natural erosion maintaining bare ground. *Hieracium pilosella* possesses several attributes that enable it to outcompete many low-growing native grass and herb plant species, especially in drier regions. These include rapid growth, vigorous vegetative reproduction, drought-tolerance, and grazing resistance. Furthermore, *Hieracium pilosella* is more efficient than native plant species in utilizing seasonal pulses of nutrients and water. The wide-

FIGURE 4 Mass flowering of *Hieracium pilosella* in *Festuca-Poa* short tussock grassland, eastern South Island, New Zealand. Photo: Stan van Uden.

spread aerial application of superphosphate fertilizer in upland areas may also have contributed to the spread of *Hieracium pilosella* in recent decades. New Zealand plants of *Hieracium pilosella* are entirely polyploid with high levels of genetic variation enabling the rapid evolution of innovative genotypes for colonizing new habitats.

The future of *Hieracium pilosella* in indigenous grasslands probably depends on land-use goals. In tall-tussock grassland, in the absence of mammalian grazing, natural vegetation cover and successional processes eventually leave *Hieracium pilosella* restricted to marginal habitats. However, the outlook for induced short-tussock grassland, with or without stock or feral animal grazers, is less clear, and *Hieracium pilosella* may well monopolize these communities in the longterm, or at least until succeeded by shrubs. Active management in areas under extensive pastoralism involves grazing to reduce flowering and applications of fertilizer to improve the competitive ability of associated grasses and legumes. A biological control program in New Zealand is currently underway, and two agents collected from Europe have been utilized: a pathogenic rust fungus (*Puccinia piloselloidarum* var. *hieracii*) and a gall wasp (*Aulacidea subterminalis*), which forms galls on stolons.

E. Spartina alterniflora (Poaceae)

Estuarine habitats, characterized by waterlogged soils and regular saltwater incursions, provide a severe environment for plants and usually support a sparse halophytic flora. One kind of plant that is able to form dense monospecific stands in these environments is the grass *Spartina* (cordgrass), especially *Spartina alterniflora* and its derivatives, which have spread in estuaries in many subtropical and temperate regions of the world. The invasion of *Spartina alterniflora* is significant because of the rapid evolutionary process involved and the growing recognition of the biodiversity values of estuarine areas.

Spartina species are perennial, deep-rooted (30 cm), rhizomatous, sward-forming erect grasses that mostly occupy low- to midtidal mudflats. They are well adapted physiologically to tolerate saline conditions with special salt excretion, dilution, and restriction mechanisms. A high water-use efficiency, due to their C_4 photosynthetic system, and the unusual ability to maintain higher rates of photosynthesis than other C_4 and many C_3 species under cool temperate conditions (5–10°C), increases their environmental range.

The native geographic distribution of the genus centers on the east coast of North and South America, with fewer species on the west coast of North America, Europe, and north Africa. *Spartina alterniflora*, native to the Atlantic coast of North America, was accidentally introduced to the United Kingdom early in the 19th century, and produced rare hybrids with the local congener *Spartina maritima*, which is restricted to western Europe and Africa. The male sterile F_1 hybrid *Spartina* x *townsendii* subsequently produced, via doubling of chromosomes, a new fertile species *Spartina anglica*. It quickly became apparent that *Spartina townsendii*, and particularly *Spartina anglica*, were markedly more invasive than either of the parent species. Estuarine habitats in southern England were transformed from exposed mudflats into tall-grass meadows. In less than a century *Spartina anglica*, via seed dispersal, clonal spread, and human plantings, covered approximately 10,000 ha of intertidal salt marsh along the coast of Britain.

Early this century, estuarine stabilization to maintain commercial waterways and reclamation for industrial and agricultural development were seen as important goals of coastal management. Experience in the United Kingdom resulted in *Spartina alterniflora* and its derivatives (*Spartina* x *townsendii*, *Spartina anglica*) being widely planted in parts of North America, in regions occupied by native *Spartina* species, and in areas well outside its native range (e.g., Australia, and New Zealand) where they were similarly successful, especially *Spartina anglica*. The heightened competitive vigor and environmental range of *Spartina anglica* appears to derive from a greater genetic heterozygosity, due to hybridization and polyploidy, and expanded phenotypic plasticity. Spreading *Spartina* species appear to have fewer insect and avian herbivores in their new ranges.

Invasive *Spartina* species induce major geomorphological changes in tidal habitats by enhancing rates of sediment accumulation. Sediment accretion rates of up 17 cm yr^{-1} (average 5–10 cm) have been recorded within the tall grass sward, and in southern England *Spartina* caused elevation of mudflat surfaces by 1.8 m in 37 years. Declining tidal influence, and the accumulation of biomass (dry weight up to 7500 kg ha^{-1}), transforms salt marsh areas into nonestuarine ecosystems. Impacts on biodiversity are no less dramatic. At the pioneer phase, establishment via seed or planting produces circular clumps that expand at rates of between 3 and 7 m yr^{-1}, eventually coalescing to form extensive meadows on sand and mud substrates in the low- to midtidal range (Fig. 5). *Spartina anglica* physically displaces many native indigenous low-growing halophytes (e.g., *Selliera*, *Salicornia*, *Schoenus*, *Puccinellia*) and modifies habitats formerly occupied by a range of bird,

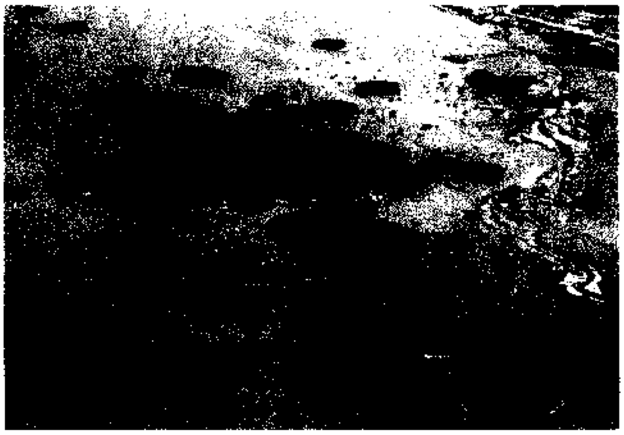


FIGURE 5 Colonizing patches of *Spartina alterniflora* on mudflats at Wilapa Bay, Washington, USA. Photo: Glen Miller.

fish, and invertebrate species. In southern New Zealand, the loss of fish and other food sources in estuaries colonized by *Spartina anglica* threatens the persistence of 34 species of foraging shorebirds in these habitats.

The hybridization of native and introduced *Spartina* species during the 19th century in the United Kingdom is presently occurring in western North America, where the recently established *Spartina alterniflora* (introduced) is genetically assimilating the common *Spartina foliosa* (native) through introgressive hybridization. Abundant *Spartina alterniflora* pollen improves seed set in *Spartina foliosa*, producing fertile hybrids that also cross with the native species. Locally, the introduced *Spartina* species and progeny could readily displace the native species and colonize lower reaches of the mudflats, currently beyond the environmental tolerance of the *Spartina foliosa*.

Because of growing appreciation of the biological values of estuaries, new plantings of *Spartina alterniflora* have largely ceased, and the species is now considered a serious weed. Extirpation of existing *Spartina* meadows has, however, been difficult, but containment and the removal of outliers has been achieved using systemic herbicides.

F. Opuntia (Cactaceae)

Cacti are prized as ornamentals for their novel form, spectacular flowers, edible parts, and medicinal uses. *Opuntia* (prickly pear) a genus of about 90 species native to North and South America, has been distributed globally over the past two centuries. Sixty species have become naturalized, and at least 15 are considered major weeds. *Opuntia* species provide some of the earliest and most spectacular examples of invasive plants in dry environments and the effectiveness of insects as biological control agents. In the major centers of *Opuntia* invasion (Australia, South Africa, and India), several species have displaced horticultural crops, reduced the grazing potential of rangelands, and threatened the conservation values of native grasslands, scrublands, and woodlands. Eradication may be unattainable, but some invasive *Opuntia* species are now at low population densities, while others are continuing to spread.

Opuntia species are fast-growing perennial succulents with thickened, often flattened, segmented cladodes, usually supporting spines. Most *Opuntia* will regenerate from seed, cladode fragments, and underground tubers. Ranging in size from low-growing shrubs to small trees, the species characteristically occupy dry habitats with seasonal water deficits. As with all cacti, *Opuntia* species have a distinctive photosynthetic system (crassulacean acid metabolism or CAM), which enables them to fix carbon at night when evaporative stress is reduced, as well as during the day, if adequate moisture is available. A high water-use efficiency, coupled with a large internal water-holding capacity and the ability to restrict water loss by tightly shutting off stomata, enhance drought tolerance. In North and South America *Opuntia* species are usually sparse and restricted regionally, and several are considered threatened in the wild. The major invasive weed species come from a range of native regions: *Opuntia stricta* (common or erect prickly pear) eastern North America, West Indies and adjacent South America; *Opuntia aurantiaca* (tiger pear, jointed cactus) temperate South America; *Opuntia ficus-indica* (sweet prickly pear, Indian fig) Central America; and *Opuntia vulgaris* (drooping prickly pear, Barbary fig) eastern South America.

Initially *Opuntia* species were distributed outside the New World as ornamental plants, but a range of uses rapidly emerged, especially as stock fodder, hedges, and as fruits or vegetables. A novel use was the production of carminic acid, a commercially important red dye, derived from the crushing of dried cochineal beetles fed a diet of cacti. *Opuntia* can provide an important source of food and income for local communities, and attempted extirpation is not universally accepted.

Introduced into Australia in the early 19th century, *Opuntia stricta* had by 1925 infested more than 25 million ha of eastern Australia and was spreading at the rate of 100 ha per hour, aided by the dissemination of vegetative and seed material by floods, humans, and feral and wild animals. Areas most susceptible were pastoral and arable land cleared amongst *Acacia* and

Casuarina woodland, although the cacti also penetrated the woodland understorey. The Australian flora has few native succulents, and none with the environmental tolerance of *Opuntia*. Approximately 10 years after initial establishment at a site, ground cover became dominated by impenetrable thickets of *Opuntia* reaching densities of 16,000 plants and a biomass of 250,000 kg, per hectare. The weed frustrated the farming ambitions of European settlers and caused the ruin of early rural economies.

Opuntia ficus-indica, a tall shrub, was established in southern Africa at least 250 years ago. By the early 20th century, infestations occupied 900,000 ha, mainly in the Eastern Cape region, occupying grassland, succulent-karoo, and the savanna biome (Fig. 6). *Opuntia aurantiaca*, a low-growing species, which may be of hybrid origin, was introduced much later (early 19th century), spreading originally from garden plantings around Cape Town. A total infested area of about 400 ha in the 1890s has grown this century to around 830,000 ha, mainly in eastern parts of South Africa. It is replacing important pastoral plants in grasslands and savanna, injuring domestic and feral animals, and degrading natural rangelands. A single cladode can produce up to 145 new cladodes over a 200-day growing period, and the potential for vegetative spread is enormous. Currently *Opuntia aurantiaca* is considered to be South Africa's most expensive weed. *Opuntia vulgare* has also been an important weed in the latter part of the twentieth century, mainly in western coastal areas.

In Australia and South Africa, the control of *Opuntia* assumed top priority for land management agencies early in the 20th century, when various herbicidal and mechanical methods were attempted. Poisoning, using a combination of arsenic pentoxide and sulfuric acid, and mechanical cutting were successfully used to control small infestations of *Opuntia* in open lands but were very expensive and achieved little at a regional scale. However, spectacular success has been achieved using plant-sucking cochineal insects *Dactylopius* species and the cladode-eating larva of the moth *Cactoblastis cactorum*. Indications of the potential of insects as biocontrol agents first appeared in 1795, when the accidental introduction of *Dactylopius ceylonicus* resulted in widespread death of *Opuntia vulgaris* in India. Subsequently, deliberate introductions of several *Dactylopius* species have drastically reduced *Opuntia* weed infestations in southern Africa, and *Cactoblastis* has been similarly used to diminish population densities of *Opuntia stricta* in Australia. Overall, biological control agents have lowered population densities of *Opuntia* by 90%, especially in drier climates. *Opuntia* remains a widespread and invasive weed in these countries, but the economic impact on agricultural land uses has diminished hugely, especially in Australia.

FIGURE 6 *Opuntia ficus-indica* invading rangeland in the Eastern Cape Province of South Africa. Photo: John Hoffmann.

G. Fallopia japonica (Polygonaceae)

The human passion for gardens increases biodiversity in urban areas, often creating large source populations, which initiate the effective dispersal of introduced plants into adjoining landscapes. One garden ornamental that has become an invasive weed is *Fallopia japonica* (Asiatic knotweed), a large perennial herb native to the Far East. In the 20th century the species has infiltrated much of central and western Europe, North America, and several southern temperate countries, where it excludes native plant species in artificial and highly disturbed habitats, riparian areas, and in open woodlands. In the United Kingdom, *Fallopia japonica* is presently the tallest and most aggressive common herbaceous species.

Fallopia japonica is a rhizomatous, clump-forming, perennial native to China, Japan, Korea, and Taiwan. In these countries it is an early successional species, growing to less than 2 m tall, colonizing volcanic debris and disturbed riparian habitats, under a range of edaphic conditions, from lowland to subalpine. *Fallopia japonica* establishes on new sites by seed in its native range and persists to develop a large biomass (12 tonnes ha^{-1}), deep litter layer, and a rhizome system that extends for up to 20 m beyond the shoots.

Taken from Japan to the United Kingdom in 1825, for planting in gardens, *Fallopia japonica* was subsequently established in other northern (e.g., North America) and southern hemisphere (e.g., New Zealand)

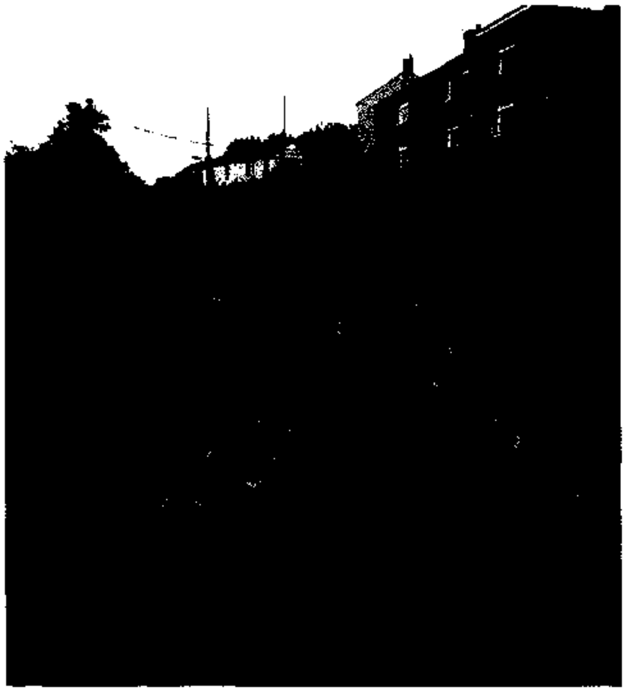


FIGURE 7 Complete cover of *Fallopia japonica* in urban wasteland, Swansea Valley, United Kingdom. Photo: Simon Fowler. See also color insert, Volume 1.

countries late in the 19th century. Although mostly found on heavily modified and disturbed sites within urban environments, it has also spread in natural habitats, especially along waterways. In the United Kingdom it is presently found in over half the 10-km grid squares, being most abundant in southern and western regions (Fig. 7). In Wales, for example, *Fallopia japonica* occupies 84% of river systems with average flow rates greater than 2.3 $m^3 s^{-1}$. In North America it has become naturalized in most mesic western coastal regions, north to Alaska, all of the northeastern United States, and parts of central and southern United States. In western Pennsylvania, stands extending for hundreds of hectares cover wetlands, stream sides, and adjoining hill slopes.

Around waterways, *Fallopia japonica* is capable of forming dense thickets up to 500 m across, overtopping and supplanting native herbs, shrubs, and small trees; impeding water runoff and increasing flooding, and modifying nutrient availability and cycling patterns through sequestration in a large standing biomass. Early emergence and fast growth of new shoots and flowering stems, usually completed before midsummer, create a dense overstory, and this, together with prodigious annual litter production, restricts the presence of other species, except where regular flooding removes surface litter, and the shade from tall trees reduces its vigor.

In its native range, *Fallopia japonica* bears female and hermaphrodite flowers on separate individuals. Elsewhere the species is male-infertile and, although seeds are produced from hybridization with other introduced congeners, spread via successful sexual reproduction is rare. However, spread of rhizome fragments is extremely potent, and viable plants can establish from pieces as small as 7 g, and from burial at depths of up to 1 m. Stem fragments are also able to produce new shoots. In most instances vegetative dispersal is achieved by water currents and the movement by humans of soil and gravel contaminated with rhizome pieces.

The current geographic range in Europe extends north to 63°N and is broadly correlated with degree-days, absolute minimum temperatures (>-30.2°C), and annual rainfall (>500 mm). The shoots are vulnerable to late frost and summer droughts, and these factors may constrain the large-scale distribution of *Fallopia japonica*, especially in continental climates. Tolerant of a broad range of soil conditions, *Fallopia japonica* is strongly light-demanding and rarely enters forests.

Once established, *Fallopia japonica* is difficult to remove. In Europe and North America young shoots of *Fallopia japonica* are occasionally eaten by stock, and this may reduce the development of clumps and slow the rate of spread. Where accessible, regular cutting or spraying with herbicides can achieve control, but follow-up treatments over several years are needed to weaken and eventually kill the rhizomes.

H. Lantana camara (Verbanaceae)

International trading in plants and the intensive search for novel cultivars for horticulture has produced new hybrids with high invasive capacity and given them near-global distributions and access to a great range of habitats. *Lantana camara* (lantana, white sage, wild sage, tick berry), a perennial shrub with brightly colored flowers, was taken from Brazil to Europe in the middle of the 17th century as a hothouse plant. During the next several hundred years, extensive hybridization and propagation of different varieties, augmented by new collections from the native region, produced a proliferation of forms. Commonly grown in gardens, *Lantana camara* has now become a major weed worldwide in tropical, subtropical, and warm temperate regions. It invades pastures, plantation crops, and a range of disturbed-open natural and modified scrubland, woodland, and forest communities, displacing the indigenous biota and limiting public access and use. Because of its broad distribution, invasive ability in both agricultural and natural ecosystems, and local persistence, *Lantana camara* is considered one of the world's top ten weeds.

Lantana camara is an aromatic shrub with distinctive four-angled stems, often armed with recurved prickles. It has multicolored, insect pollinated, bisexual, potentially self-fertile flowers that produce purplish-black drupes by assorted sexual, semisexual, and apomictic mechanisms. Native to parts of Central and South America, *Lantana camara* is actually a complex of species and varietal forms of mixed hybrid origin. It is a genetically diverse group, with ploidy levels ranging from diploid to hexaploid. The most weedy variety (*Lantana camara* var. *aculeata*) is tetraploid. The growth form (e.g., ground creeper, prostrate shrub, tall shrub, liane) of *Lantana camara* is equally plastic, varying with light and soil conditions, but with support plants can reach 15 m into the canopy.

The species currently grows in many countries between 45°N and 45°S, but appears to be limited to minimum mean monthly temperatures above 5°C, rainfall in excess of 650 mm per annum, and nonsaline soils. It is a serious weed in the Caribbean, in eastern Africa, South Africa, India, Madagascar, Australia, and the Pacific Islands.

Lantana camara is an important weed of plantation crops. In India, it has taken over land used for tea and sugarcane production, causing the displacement of whole villages, and in other countries it is a major nuisance in cotton, coffee, coconut, oil palm, bananas, pineapple, rubber, and rice crops. In continental countries such as Australia, India, and South Africa, it is natural grasslands that have been most extensively modified by *Lantana camara*. In eastern Australia alone approximately 4 million ha of grassland have become covered, eliminating both indigenous species and pastoral agriculture. Remnants of semideciduous forest are also invaded via edges and natural and human disturbances of the canopy. In South Africa, 44% of 1-km grid squares were infected with *Lantana camara* in a survey of 14 major forest reserves.

The spread of *Lantana camara* on remote volcanic islands with distinctive biotas has been of international concern and has usually been associated with European settlement and vegetation clearance via fire, roads, and grazing mammals (Fig. 8). On the Galapagos Islands, famous for their biodiveristy and role in the origin of evolutionary theory, *Lantana camara* endangers both rare plants and animals. It is considered to have caused the extinction of one plant species and threatens the demise of at least eight others. The weed is also closing in on the last remaining colony of the darkrumped petrel (*Pterodroma phaeopygia*), and will, if not managed, seal off bird access to burrows. In the Indian Ocean, *Lantana camara* is one of several shrub-creepers invading and depleting native forests on La Réunion Island. Initially colonizing natural gaps, it subsequently ascends to smother and weaken canopy trees, making them more susceptible to damage during cyclones. It is predicted that *Lantana camara* will increase dominance with successive cyclones and eventually displace the forest cover. Thirteen years after its introduction to the Hawaiian Islands in 1858, *Lantana camara* had become naturalized on all the islands and is considered to be the number one problem plant. It currently occupies more than 160,000 ha, mainly in coastal and lowland areas, covering cropland, shrub land, savanna, and destroying opened up forest.


FIGURE 8 Flowering *Lantana camara*, a major threat to indigenous biodiversity on oceanic islands. Photo: Simon Fowler.

Apart from a broad environmental tolerance, the species has numerous traits that facilitate its dispersal, penetration, dominance, and persistence in different ecosystems. In tropical climates flowers and seeds may be produced all year, and fruits are eaten and widely dispersed by a range of birds and mammals, including some similarly invasive introduced species (e.g., Indian mynah, *Acridotheres tristis*). Most disturbed vegetation types are vulnerable to invasion by *Lantana camara*. Grassland, shrub land, and woodlands, from coastal to mountain areas, are transformed into dense stands of *Lantana camara* as it smothers herbs and other shrubs. Continuous closed forest is less susceptible as *Lantana* has greatly reduced vigor in deep shade. However, plants can penetrate forest from around the margins, readily establish in gaps created by natural treefall, and prevent the regeneration of canopy trees. Plants are flammable, even when green, and facilitate the spread of forest fires. Foliage and seeds are toxic for many grazing mammals and the thick, prickly stands of *Lantana camara* are virtually impenetrable. Thickets also provide habitats for vermin and disease vectors (e.g., tsetse fly).

Techniques for effectively controlling *Lantana camara* are slowly emerging, although most depend on the rapid establishment of vigorous crops or pastures following initial suppression of the weed. Mechanical methods are generally less successful, because *Lantana camara* generally resprouts from plant fragments, basal shoots, and roots after fire, cutting, or digging. Herbicides, especially those based on phenoxy or benzoic acid or pyridine, have variable success depending on climate, season, and growth form. Biological control appears to be sometimes effective, especially in drier areas, utilizing a mixture of seed- and foliage-feeding insects.

I. Mimosa pigra (Fabaceae)

Nitrogen-fixing shrub legumes provide some of the most aggressive and disruptive invasive plant species in many parts of the world. One of the most spectacular plant invasions this century has been the spread of *Mimosa pigra* (giant sensitive plant, zaraz, dormilona) in tropical wetlands in parts of Asia, Africa, and Australia. In less than a century, 450 sq km of natural habitat associated with flood plain areas and rivers around Darwin, northern Australia, have been transformed into dense stands of *Mimosa* (Fig. 9). The species endangers the conservation and use of natural wetland ecosystems in tropical regions worldwide.

FIGURE 9 Dense stands of *Mimosa pigra* during the wet season, northern Australia. Photo: CSIRO Entomology.

Mimosa pigra is a prickly shrub native to Central and South America where it forms shrub lands up to 5 m tall in areas with seasonally high humidity. Its current pantropical distribution reflects human movement of plants since the 16th century, most likely because of fascination with the touch-sensitive rapid folding of the foliage. *Mimosa pigra* has naturalized beyond its native range in Asia (e.g., Thailand, Malaysia), America (e.g., Costa Rica, Brazil), Africa (e.g., Namibia, South Africa), and Oceania (e.g., Australia, New Guinea). Climatically, *Mimosa pigra* favors the seasonally dry tropical zone, with an annual rainfall of between 750 and 2250 mm, mean annual temperature above 17°C, and mild winters.

Characteristics that make *Mimosa* such an aggressive invasive plant species include rapid growth rate (10 mm per day), rapid maturation (germination to first flowering within 6 months), potentially autogamous, abundant seed production (9000 seeds per m^2 annually), a large, long-lived (>10 years) seed bank in soil, and an effective dispersal system (flotation of clusters of capsules and via attachment to animals). Spines on the stem and leaf rachis deter most mammalian grazers and, outside its native range, there appear to be few invertebrate or fungal attackers. In many habitats, *Mimosa pigra* has the ability to completely dominate a site, forming impenetrable, monospecific shrub stands up to 6 m tall.

Rates and pattern of invasion have been best documented for northern Australia. Movement away from the entry point in Darwin has taken more than 50 years, but spread since the 1970s has been rapid. Within a river system stands expand on average at the rate of 76 m yr^{-1}, resulting in a doubling time of 1.4 years. Peripheral expansion is fastest following higher than average rainfall during the wet season, suggesting the importance of water for both the dispersal of seeds (in the wet season) and the survival of seedlings (through the dry season). An example of this spectacular rate of local spread has been given for a site of known history where a stand grew from a few individuals to one covering an area of 60,000 ha in approximately 10 years. At a larger scale, the doubling time for new infestations is slower (6.7 years), due to habitat heterogeneity.

The role of introduced grazing animals in the invasion of *Mimosa pigra* is debated. In northern Australia, the feral asiatic water buffalo (*Bubalus bubalis*) was thought by pastoralists to limit the spread of *Mimosa*, because of its conspicuous expansion since the lowering of buffalo densities as part of a disease (brucellosis and tuberculosis) eradication program in the mid-1980s. However, in the two decades prior to this, both buffalo numbers and *Mimosa pigra* infestations increased in parallel, and recent evidence shows that the most readily invaded habitats (wetland margins) are those most disturbed by large grazers, lacking tall trees, and with long periods of inundation.

Mimosa pigra drastically alters the composition and use of natural ecosystems. In Australia, it has supplanted native sedge land and grassland communities on flood plains and has invaded and displaced adjoining *Melaleuca*, *Eucalyptus*, and *Pandanus* woodland. The tall cover of *Mimosa* decreases native biodiversity and threatens the rich wildlife associated with open habitats. Ecosystem modifications caused by *Mimosa pigra* have reduced native resources accessible for traditional aboriginal use, pastoral grazing, and ecotourism ventures. The enormous scale and potential level of impact of *Mimosa pigra* is highlighted by estimates provided for Kakadu National Park (13,000 km^2), a World Heritage Area containing a full range of natural habitats within 150 km of the coast. Over 80% of the park is vulnerable to invasion (29% seasonally flooded areas susceptible to complete displacement; 54% largely open forest and woodland, which could have *Mimosa pigra* as a common element). Only 17% (extreme edaphic sites) of the area appears beyond the ecological tolerance of *Mimosa*

The management and control of *Mimosa pigra* in natural ecosystems present a major challenge. It is susceptible to mechanical removal, herbicides, fire, and competition from grasses, and all are being used, often in combination, to eradicate new infestations and control the expansion in Australia. However, to be effective, treatments need to be repeated at regular intervals for at least a decade due to resprouting from damaged stumps and recolonization via the seed bank (10^4 seeds m^2). Currently an extensive biological control program is being undertaken testing several insects and pathogens.

J. Pinus pinaster (Pinaceae)

Members of the Pinaceae, notably *Abies*, *Larix*, *Picea*, and *Pinus*, are the most widely planted exotic tree species in the world and form the basis of production forestry in many countries. The evolutionary history of the family is centered on the northern hemisphere, but in the 20th century the geographic range of northern conifer species has been greatly extended to many southern hemisphere regions where they have been evaluated and used for forestry, amenity planting, reforestation, and erosion control. Translocation of *Pinus* species in particular has occurred on a grand scale, initiated by European settlers. In South Africa, for example, more than 80 of the 111 species worldwide have been planted since the late 17th century. Outside their native range, at least 20 species have established away from plantations into adjoining grassland, shrub land, and woodland communities. Because of their size, growth rate, and fecundity, pines have been responsible for generating widespread concern internationally about the potential impacts of introduced plant species. *Pinus pinaster* (cluster pine, maritime pine), native to the Mediterranean region, typifies the invasion history and types of impacts that pines can have, particularly in nonforest ecosystems.

Naturally occurring in the Mediterranean Basin from Portugal to Algeria, *Pinus pinaster* is a two-needled, tall (40 m), open-canopied pine species frequently growing on sandy soils and dunes among oaks and heathland shrubs at low to moderate altitudes, but extending up to 1500 m. Valued for resin potential and for erosion control, *Pinus pinaster* has been planted widely in South America, Australasia, and South Africa, where it has invaded a range of natural and modified communities. Plantations have also been established in Europe, within its native range, and these are similarly spreading.

The ecology and spread of *Pinus pinaster* has been best documented in South Africa, where it was introduced in the late 17th century and subsequently widely planted for timber and drift sand stabilization. Two hundred years later *Pinus pinaster* has invaded 3256 km^2 of natural vegetation (Fig. 10) and is one of several pine species (*Pinus halepensis*, *Pinus radiata*) that threatens the conservation of one of world's biodiversity gems, the fynbos biome. Extending for 71,337 km^2 in the Cape Floristic region, the fynbos biome comprises complex, species-rich (7300 vascular plants), fire-adapted,

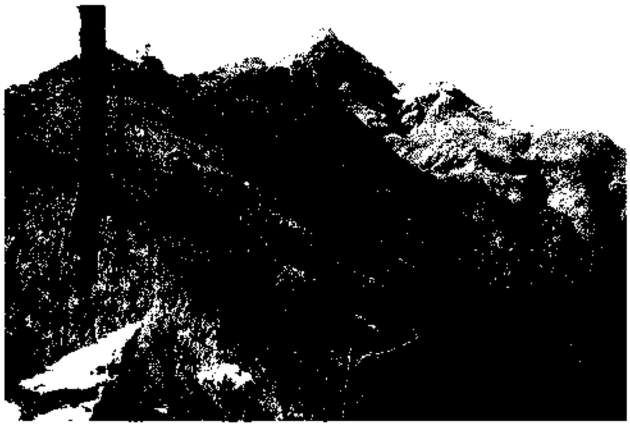


FIGURE 10 Thousands of hectares of *Pinus pinaster*, after 100 years of spread in the Riviersonderend Mountains near Genadendal, Western Cape Province, South Africa. Photo: David Richardson.

mainly shrub-dominated ecosystems with a high proportion (80%) of local endemics. The pines have spread peripherally around source plantations and also via long-distance (up 5 km) wind dispersed seed. These outliers gradually coalesce to form extensive stands in which *Pinus pinaster* is more than 25% of the cover. On the Cape Peninsula near Cape Town, pines are the most common introduced plant species in 10 of the 15 vegetation types represented at the Cape, with *Pinus pinaster* dominant over 560 ha or 1.1% of the total area.

The success of *Pinus pinaster* and other invading pines is due to a combination of factors including particular species traits, natural features of the fynbos, and various human activities. Weedy pines, in this ecological context, are distinguished from nonspreading species by their resilience to fire (e.g., short period to reproductive maturity, above-average levels of serotiny) and dispersibility (small seeds with low seed-wing loading). The fynbos vegetation is fire prone, and natural fires at intervals of a decade or more maintain a mosaic of shrub land and gully forest, depending on the timing and intensity of the burn. Fire frequencies have increased with greater human activity in the region, creating ideal invasion opportunities for fire-adapted, fast-growing, and early-maturing introduced species. Plants derived from abundant seed sources blown in from established plantations have penetrated the fynbos after fire, increasing (by up to 300%) the available flammable biomass, and gradually consolidating dominance with successive fire cycles.

In the absence of management, *Pinus pinaster* and the other pines would eventually transform much of the diverse fynbos vegetation into an introduced conifer forest. The shift in life-form dominance threatens some rare native plant species and will suppress many other light-demanding plants, although they may provide new habitats for arboreal native birds. The large increase in stand biomass caused by pine invasion has increased interception and transpirational losses resulting in decrease stream flow (by between 30 and 70%) for human consumption from pine-dominated catchments. The pines have induced a new vegetation structure and system, mainly as a result of increasing the intensity of burning, and few native species can cope with the new disturbance regime.

Management for the conservation of the fynbos biome must include controlling the invasive pines. Where plantations adjoin watersheds, massive clearing operations are financially justifiable in view of the major impact of invading pines on water delivery from catchments.

Options currently being undertaken involve prescribed burning at intervals of 12 to 15 years, in conjunction with mechanical clearing of pine stands and outliers. Pines are felled 12 to 18 months before burning to allow time for seeds to be released from serotinous cones, and seedlings to establish. These are subsequently killed by the fire. Biological control approaches using seed-attacking insects are being explored for *Pinus pinaster* and *Pinus halepensis* as long-term solutions for effectively managing pines in fynbos landscapes.

III. PREDICTING AND EVALUATING IMPACTS

Most countries want to limit the influx of potentially invasive plant species and to identify and contain those that already exist within the large reservoir of introduced plants confined to cultivated areas. The commonest approach to improving biosecurity at the border and enabling recognition of new weeds at a very early stage of invasion when control is both cost-effective and practicable is to apply some form of weed risk assessment analyses. These invariably include evaluations of potential impact on human activities (e.g., agricultural production, animal health) and natural environments. Information is commonly derived from weed databases on the behavior of the plant species elsewhere in the world or from assumptions about the relation between plant traits and species impact. In general, features that enable a species to either capture or produce an important resource will have the greatest effect on plant community composition and structure and will initiate major changes in ecosystem processes and functions. Key traits vary depending on the local environment and the type of land management, but those that either assist species to acquire an unequal share of key resources (e.g., relative growth rate, height), foster new disturbance regimes (e.g., fire), enhance local persistence (e.g., long-lived soil seed bank), or create a major increase in resource availability (e.g., nitrogen-fixing rhizobia) are likely to be the most important in any risk analysis. However, stochastic processes (e.g., change in land use), complexity of multiple interactions in natural ecosystems (e.g., presence of seed dispersers), and the increasing prevalence of introduced plant species lacking a weed history elsewhere collectively contribute to making weed impact prediction a major challenge. Methods for evaluating comparative benefits (e.g., food source for humans and native animals, erosion control) and costs (e.g., loss of crop production, reduced native biodiversity) of introduced invasive weed species are only slowly emerging due to difficulties in assessing qualitative and quantitative elements of these categories in simple economic terms.

Acknowledgments

This article benefited from comments from Stephen Brown, Garry Cook, Curt Daehler, Penny Edwards, Barrie Forrest, Peter Grubb, Simon Fowler, Mic Julien, Richard Mack, Wendy Nelson, Trevor Partridge, Ian Payton, Dave Richardson, Alan Rose, Leslie Seiger, and Peter Williams.

See Also the Following Articles

FIRES, ECOLOGICAL EFFECTS OF • INTRODUCED SPECIES, EFFECT AND DISTRIBUTION • MIGRATION • PLANT BIODIVERSITY, OVERVIEW • PLANT INVASIONS

Bibliography

Brock, J. H., Wade, P., Pysek, P., and Green, D. (Eds.) (1997). *Plant Invasions: Studies from North America and Europe.* Backhuys, Leiden.

Cowling, R. M., Richardson, D. M., and Pierce, S. M. (Eds.) (1997). *Vegetation of Southern Africa.* Cambridge University Press, Cambridge.

Cronk, Q. C. B., and Fuller, J. L. (1995). *Plant Invaders: The Threat to Natural Ecosystems.* Chapman and Hall, London.

Groves, R. H., Shepherd, R. C. H., and Richardson, R. G. (Eds.) (1995, 1998). *The Biology of Australian Weeds.* R. G. Frankton and F. J. Richardson.

Luken, J. O., and Thieret, J. W. (Eds.) (1997). *Assessment and Management of Plant Invasions.* Springer, New York.

Pysek, P., Prach, K., Rejmanek, M., and Wade, M. (Eds.) (1995). *Plant Invasions: General Aspects and Special Problems.* SPB Academic Publishing, Amsterdam.

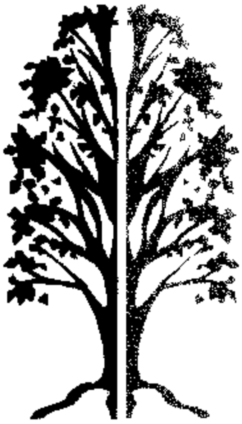
Richardson, D. M. (Ed.) (1998). *Ecology and Biogeography of Pinus.* Cambridge University Press, Cambridge.

Simberloff, D., Schmitz, D. C., and Brown, T. C. (Eds.) (1997). *Strangers in Paradise: Impact and Management of Non-Indigenous Species in Florida.* Island Press, Washington, D.C.

Starfinger, U., Edwards, K., Kowarik, I., and Williamson, M. (Eds.) (1998). *Plant Invasions: Ecological Mechanisms and Human Responses.* Backhuys, Leiden.

Vitousek, P. M., D'Antonio, C. M., Loope, L. L., Rejmanek, M., and Westbrooks, R. (1997). Introduced species: A significant component of human-caused global change. *New Zealand Journal of Ecology* 21, 1–16.

Walker, D. I., and Kendrick, G. A. (1998). Threats to macroalgal diversity: Marine habitat destruction and fragmentation, pollution and introduced species. *Botanica Marina* 41, 105–112.



INTRODUCED SPECIES, EFFECTS AND DISTRIBUTION OF

Daniel Simberloff
University of Tennessee

GLOSSARY

biological control Introduction of a natural enemy (parasite, predator, herbivore, or pathogen) of an undesirable species, usually itself introduced.

coevolution Evolved mutual adaptations between species, in which each species influences the evolution of the other.

hybridization Mating between individuals of two different populations, usually classified as separate species. Mixing of their genomes is not required.

introgression Gene flow between populations whose individuals hybridize, achieved when hybrids backcross to one or both parental populations.

island biogeography, theory of A theory that the number of species in the biota of each island (or island-like habitat) is in dynamic equilibrium, with species frequently going extinct on the island and new species frequently arriving.

I. INTRODUCTION

What constitutes an effect of an introduced species (often called "impact") has never been formally defined. One common use of the term is for interactions with native species (e.g., predation), as well as consequences of these interactions for population, community, and ecosystem structure and function of the system to which the species was introduced. Another frequent referent for this term is the economic cost generated by an introduced species (including loss of goods or services and costs of control). A range of less frequent meanings of "effect" will be clear as effects are described. Nor is the meaning of "introduced species" clear-cut. It is widely used for any species introduced to a new region through human agency, whether deliberately or accidentally. Some authorities wish to restrict "introduced" to species deliberately imported by humans, and describe species that arrive with human assistance but that are not deliberately imported as "immigrants." In this restrictive definition, introduced species plus immigrants compose the category "nonindigenous species." A few authors also include as introduced species those that arrive on their own in a location far from their original homes—for example, the Old World cattle egret (*Bubulcus ibis*) in Florida. Finally, an introduced species is often defined legally as one not native to a particular nation; a species carried by humans from one location to another within a nation would not be introduced. Many other common terms—alien, exotic species, bioinvaders—are less precisely used. In this article, introduced species will be those that arrive in a new region with human assistance, deliberately or inadvertently, whether or not the new region is within the national borders of the original range.

The effects of introduced species are numerous, often subtle, and idiosyncratic. Because intensive study of invasions by introduced species is quite recent, many statements about their effects are based on intuition rather than empirical data. Experimental data on effects of introduced species are rare. Intuition is sometimes incorrect. Thus, the decline of the native otter (*Lutra lutra*) in the United Kingdom was long attributed to the introduction of the American mink (*Mustela vison*), whereas more recent research exonerated the mink and implicated organochlorine pesticides. Conversely, the destruction of the forest birds of Guam was at first attributed to pesticide poisoning before the introduced brown tree snake (*Boiga irregularis*) was proven to be the cause.

Another complication in predicting and determining effects of introduced species is that, after introduction, there is frequently a time lag before they spread and begin to exert an impact. For example, Brazilian pepper (*Schinus terebinthifolius*), which now infests about 300,000 hectares (ha) in south Florida, was an innocuous, restricted species for over 50 years. Sometimes the causes of these lags seem evident (e.g., delay in the arrival of an obligatory pollinator for a plant); other times they are mysterious. The opposite phenomenon—dramatic increase in numbers and impact by invaders, followed by a decline independent of directed human actions—is also occasionally observed. For instance, on several Pacific islands that were massively invaded by the terrestrial giant African snail (*Achatina fulica*), a rather abrupt decline in density was observed. Again, causes sometimes suggest themselves (e.g., a pathogenic disease), other times they do not.

Despite uncertainties surrounding the effects of introduced species, a recent burst of research shows that they are often dramatic and, in sum, constitute a major, ongoing global change. They are widely recognized as the second greatest cause of species endangerment and extinction (after habitat change, with which they often interact). A recent estimate by D. Pimentel and others suggests that the annual cost in the United States alone is greater than $130 billion. This is not to say that most invaders have dramatic impacts. They do not; most do not even survive. M. Williamson's widely cited tens rule (about 10% of introduced species establish without human assistance, and about 10% of those have effects large enough to be viewed as pests) is certainly too simplistic, but it accurately depicts the lack of substantial impact by most invaders.

II. THE GEOGRAPHY AND MAGNITUDE OF INVASION BY INTRODUCED SPECIES

A. How Many Are There, Where Do They Come From, and Where Do They Go?

No comprehensive list of introduced species exists for most taxa in most locations; indeed, there is often no list of native species. However, for well-studied groups, some figures are impressive. In Florida, for example, 27% of established plant species, 8% of insects, 29% of land snails, 24% of freshwater fishes, 22% of reptiles, 5% of birds, and 24% of land mammals are introduced. For the Hawaiian Islands, as for many islands, some of the analogous figures are even greater: almost half the plants, 25% of insects, most freshwater fishes, and 40% of birds. In some areas (e.g., Alaska) there are far fewer introduced species, but almost no regions are immune.

Although the absence of adequate quantitative data makes it difficult to describe this pattern fully, it is widely believed that Eurasian species are more likely to invade other regions than vice versa and are more likely to have large impacts. For instance, most major human pathogens and most plant pathogens that have had global impacts originated in Eurasia. Similarly, more Eurasian insects, vertebrates, and plants have invaded other regions than vice versa. The reasons for this imbalance are obscure. Some authors have argued that Eurasian species have an innate superiority, generated either by the larger numbers of species evolving greater competitive ability on the larger landmass, or by the happenstance that Eurasian species were highly coevolved—plants, pathogens, and animals (including especially grazers and humans)—and overwhelmed native species by their joint action. For example, the structure and behavior of Eurasian hooved livestock were devastating to native tussock grasses in the North American prairie but were favorable for Eurasian turfgrasses, which now dominate vast regions.

However, even if Eurasian species, singly or in groups, were not innately superior, one might have expected a preponderance of them among introductions, because the opportunities for such species to reach other regions were greater throughout most of recent history. For instance, the majority of introduced insects in the United States through the eighteenth century came from Europe in soil ballast, which was loaded in Europe and exchanged in North America for various raw materials. Similarly, as Europeans colonized other regions, they formed acclimatization societies to introduce the birds of their homelands. Except for game birds, there was little analogous movement of species in the opposite direction. Nowadays, as travel and trade are burgeoning worldwide, opportunities for introduction from any region to any other one are greatly enhanced, and one might expect habitat and climatic matching to become more important as limiting factors.

B. When Did They Get There?

The timing of introductions has depended heavily on available means of transport and patterns of travel and trade, and it has tended to increase strongly from the late eighteenth century through the present. For example, introductions of aquatic plants and animals into the Great Lakes rose steadily from one species between 1810 and 1839 to over 40 between 1960 and 1990. The advent of rapid steamship transport across oceans coincided with a dramatic increase in introductions of many taxa, as hitchhikers that could not have stayed alive over a month or more in transit were able to survive a voyage of two weeks across the Atlantic. Air travel decreased the need to survive a long transit period still further. Overlaying this dominant pattern of increasing rates of introduction with increased transport volume and decreased transit time are idiosyncrasies associated with particular taxa and regions. For example, beginning around 1920, there was a decrease in the rate at which foreign insect species, especially herbivorous species, were introduced to the United States; this downturn coincided with the enforcement of plant quarantine laws. At the same time, a dramatic increase of introduction of wasp species reflected greatly increased biological control efforts, especially the use of parasitic wasps.

C. Distribution of Introduced Species among Habitats

Habitats modified and/or routinely disturbed by humans generally house more introduced species and larger populations of introduced species than do pristine habitats, though even the latter are occasionally invaded. The reason for this pattern has been hotly debated. In general, the very fact of human disturbance renders a habitat less suitable to the native species that had evolved in the original habitat. On the other hand, some species, often associated with humans, are superbly adapted to habitats that humans create. Thus, certain introduced plants are routinely found on lists of serious weeds from many parts of the world. A common claim that disturbed habitats are invasible because they are species-poor is incorrect. Many habitats that are relatively undisturbed by humans, such as salt marshes and mountaintops, have few native species, but they also have few introduced species, and for the same reason—they are biologically difficult environments and few species of any kind have the physiological adaptations that permit them to thrive there. Conversely, enormously diverse tropical communities often have as many invaders (with as large impacts) as less species-rich temperate analogs. Nor does periodic disturbance per se automatically lead to invasiveness. For example, the intact upland pine forests of north Florida are far less stricken by invaders than many other habitats in the state, yet they frequently burn naturally because of lightning strikes. If anything, this disturbance favors the many native species that are adapted to a fire regime over a plethora of potential invaders that are not.

Most well-known examples of introduced species with enormous impacts come from either terrestrial or freshwater habitats. Marine habitats are poorly represented in the invasion literature. Freshwater habitats—lakes and rivers—can often be seen as habitat islands surrounded by land, and thus their communities are believed to be inherently invasible for the same reasons that islands appear to be particularly invasible, as will be discussed in the following section. Certainly, many lake and river communities have evolved without certain types of organisms, such as large, predatory or folivorous fishes, and thus their species are particularly prone to damage from introduced species of these particular sorts. Whether marine habitats have suffered less impact from introduced species is highly questionable. Terrestrial habitats are immediately visible, so the effects of introduced species are surely more likely to be detected there than in the sea, just as the general ecology of most marine habitats is poorly studied and understood relative to that of most terrestrial habitats. Additionally, it is now widely recognized that many marine species that had been viewed as native over wide regions are in fact of unknown provenance—they are "cryptogenic." In most places they may well have been accidentally introduced by humans. Finally, a number of marine invasions have been observed lately that vie in apparent impact with the most heralded terrestrial and freshwater invasions. The tropical alga *Caulerpa taxifolia*, purged from the public aquarium in Monaco in the early 1980s, has spread to cover 5000 ha of nearshore habitats off the coasts of Spain, France, Italy, Monaco, and Croatia. It has particularly overgrown and replaced meadows of marine grasses such as *Posidonia oceanica,* the nursery and home of many fishes and invertebrates. In the Black Sea, the western Atlantic ctenophore *Mnemiopsis leidyi* was first recorded in 1982. By 1988, its biomass exceeded that of all other zooplankton combined, and, because this comb jelly is a predator, its impact on other animals, including young fishes, must be enormous. It is widely credited with the dramatic crash of several commercial fisheries.

D. Island Vulnerability

Islands are notoriously prone to ecological damage by introduced species. Many island bird species and subspecies have been eliminated by introduced rats, feral cats, and other introduced carnivores. In the West Indies, the small Indian mongoose (*Herpestes javanicus*) has extinguished several species of endemic snakes and lizards. Introduced grazers have caused many plant extinctions on islands. The introduced carnivorous rosy wolf snail, *Euglandina rosea,* has eliminated at least seven species of terrestrial snails on the island of Moorea alone, as well as many populations of endemic tree snails in the Hawaiian Islands. Worldwide, about 1.6 times as many mammal species and three times as many bird species have been successfully introduced to islands than to mainland areas.

The reasons for the great number of introduced species on many islands, and their apparently great effects on many of them, are probably threefold. First, there are generally more deliberate attempts to introduce species to islands. As mainland inhabitants colonized islands, they saw them as biotically depauperate and attempted to introduce species—often those from home—to fill the perceived gap. Second, many of the islands that have been most assaulted by introduced species were massively modified largely independently of introductions. For example, almost all the lowland areas of the Hawaiian Islands, Mascarenes, and many of the West Indies were cleared of native species for agricultural and housing purposes. That introduced species largely replaced them does not necessarily mean that the native species were in some sense weaker than the invaders; rather, the habitats to which the native species were adapted were simply removed. Because of the smaller size of islands than mainland, habitat destruction on islands was generally much less likely to leave refugia in which the native species persisted. Third, the biotae of oceanic islands often evolved without entire groups of organisms, and so lacked adaptations to survive their impacts once they were introduced. Thus, most oceanic islands lack native predatory mammals, so birds evolved to nest on the ground and were highly susceptible to the invasion of such species. Similarly, native browsing and grazing herbivorous mammals are lacking from islands, so the native plants have not evolved traits to survive herbivory when the mammals are introduced.

III. DIRECT EFFECTS

A. Habitat Modification

The most widespread impacts of invaders are usually due to modification of the habitat, which affects many species together. Almost all introduced species believed to have had major impacts on entire ecosystems or landscapes have done so by modifying habitats.

1. Grazing and Rooting

Goats were introduced to St. Helena in 1513 and built up enormous herds. These are believed to have extinguished at least half of some 100 endemic plant species before botanists had even visited the island. This change, of course, is permanent. On the island of South Georgia, grazing by introduced reindeer (*Rangifer tarandus*) has heavily degraded tussock grassland, mesic meadow, dwarf-shrub sward, and dry meadow communities. These changes have facilitated erosion and the invasion of exotic plants. In many places, these changes resulted in a totally new plant community. Recovery of certain of the plant communities, and the insects that inhabit them, may take a long time, but exclosure experiments show that few of the changes are irreversible. European wild boar (*Sus scrofa*), feral domestic hogs, and hybrids between them have wreaked havoc in both bottomland and upland forests in the United States. For example, in the Great Smoky Mountains National Park, they "root" primarily in high-elevation deciduous forests in the summer, reducing understory cover and number of species. By selective feeding, they can locally extinguish plant species with starchy bulbs, tubers, and rhizomes. They greatly modify soil characteristics by thinning the forest litter, mixing organic and mineral layers, and accelerating mineral leaching.

2. Shading and Overgrowth

South American water hyacinth (*Eichhornia crassipes*) blankets many nearshore areas of Lake Victoria in East Africa, some lakes and rivers of the U.S. Southeast, and water bodies in many other regions. It blocks light and smothers beds of native submersed vegetation. Decaying water hyacinth can deposit over 1000 metric tons (wet weight) of detritus per hectare per year, which in turn can heavily modify water chemistry, such as nutrient concentrations. When dissolved oxygen drops, it affects many animals and the entire biotic community changes. Terrestrial plants such as Asian kudzu (*Pueraria montana*) in the Southeast can similarly cover existing vegetation and eliminate it by shading, as can marine plants such as the tropical alga *Caulerpa taxifolia* in the Mediterranean.

3. Modified Fire Regime

In much of the American West and Hawaii, Old World grasses such as cheatgrass (*Bromus tectorum*) increase the frequency and intensity of fires, greatly harming native plants and the animals that use them. Similarly, the Australian melaleuca (*Melaleuca quinquenervia*) has a spongy outer bark and highly flammable foliage, and it also produces a more abundant litter than native herbaceous communities in south Florida. These features have produced an altered fire regime that has facilitated the invasion of 200,000 ha. Both tree and ground fires are now more intense and frequent, to the detriment of the native plants.

4. Modified Hydrology

Mediterranean salt cedars (*Tamarix* spp.) have invaded the U.S. Southwest. They are deeply rooted and transpire rapidly; once established, they can survive on water deep in the soil, and their transpiration is a significant pathway of water loss in arid areas. For example, at Eagle Borax Spring in Death Valley, California, within 25 years of invasion by salt cedar, the surface water of what had been a large marsh had disappeared completely, along with the majority of its associated biota. In Israel, Australian *Eucalyptus* trees have been deliberately used to drain swamps and bogs, thus eliminating the original vegetation and animals dependent on it. In addition to changed evapotranspiration rates, plants can also affect hydrological regimes by changing soil elevation. The heavy litter of melaleuca has had this impact in the Florida Everglades.

5. Modified Nutrient Regime

Many plant invaders fix nitrogen, and in nitrogen-poor areas this added increment can be detrimental to native species, which have evolved to thrive in a low-nitrogen environment, and favor other invaders. The Atlantic nitrogen-fixing shrub *Myrica faya* has invaded young, nitrogen-deficient volcanic regions of Hawaii. There are no native nitrogen-fixers, and the invader thus alters productivity, nutrient cycling, and ecological succession. Because many nonindigenous plant species in Hawaii are more successful on more fertile sites, *M. faya* may enhance the likelihood of other invasions.

B. Competition

Competition for resources is often difficult to demonstrate, but there is strong reason to believe that introduced species often use some resource so effectively that they deprive native species of it. In addition, an introduced species can also depress a native one not by eliminating some resource but by direct interference.

1. Resource Competition

Shading and water depletion, already discussed, are means by which some introduced plants outcompete native species for a crucial resource. Introduced animals have similar effects. For instance, the house gecko (*Hemidactylus frenatus*) has invaded many Pacific islands, and it depresses the insect food base locally so that some native lizard populations decline. In Great Britain, the greater foraging efficiency of the introduced American gray squirrel (*Sciurus carolinensis*) has led to the decline of the native red squirrel (*S. vulgaris*).

2. Interference Competition

The South American fire ant (*Solenopsis invicta*), which has spread throughout much of the southeastern United States, attacks individuals of many native ant species and is replacing some species, such as the "native" fire ants (perhaps themselves pre-Columbian introductions) *S. geminata* and *S. xyloni* in several habitats. In a plant analog of aggression, the African crystalline ice plant (*Mesembryanthemum crystallinum*) accumulates salt, and the salt remains in the soil when the plant decomposes. In California, this ice plant excludes native plants that cannot tolerate the salt. In both of these examples, the invader does not actually render a resource in short supply for native species; rather, it directly inhibits the native.

C. Predation

Some of the most striking effects that invaders have had on particular native species, or groups of them, entail their preying on them. The Nile perch (*Lates nilotica*), introduced to Lake Victoria, eliminated many species of endemic cichlid fishes. Introduced rats (*Rattus* spp.) on many islands have destroyed at least 37 species and subspecies of island birds worldwide. The brown tree snake and the rosy wolf snail, mentioned previously, are predators that have extinguished many native prey species. These are all dramatic examples of how an introduced predator can affect a native fauna that has not evolved with it. However, even prosaic introduced predators may have a substantial impact on native prey. For instance, there are up to 30 million feral house cats in the United States and in Wisconsin and Virginia alone they are believed to kill more than six million birds annually.

Many predators have been introduced deliberately for the biological control of previously introduced species. The introduction in 1889 of the Australian vedalia beetle (*Rodolia cardinalis*), which controlled cottony-cushion scale (*Icerya purchasi*) on California citrus, is an early example, and there are many other successes. Some predators introduced for biological control have devastated nontarget native species while exerting little if any control on populations of the target. The rosy wolf snail (*Euglandina rosea*) discussed earlier is a prime example; it was introduced to many islands to control the giant African snail. Similarly, the small Indian mongoose, previously mentioned as having extinguished native herpetofauna in the West Indies, was introduced there and on the Hawaiian Islands, Fiji, Mauritius, and other islands for the biological control of rats. On all of these islands, it preyed on native, nontarget species, sometimes to the point of endangerment or extinction. The mosquitofish (*Gambusia affinis*) from Mexico and Central America, introduced in Europe, Asia, Africa, Australia, and many islands for mosquito control, is widely seen in many regions as being at best no better than native predators at controlling mosquitoes. It often preys on other small fishes as well as invertebrates, some of which prey on mosquito larvae. For instance, in Australia, it is implicated in the extinction of several endemic fish species through predation and both interference and resource competition.

D. Herbivory

The extinction of half of the native flora of St. Helena by goats shows that herbivory by introduced species on native ones can have major effects. There are numerous other examples in a variety of habitats. Chinese grass carp (*Ctenopharyngodon idella*) have been introduced to Europe, Africa, and North America for the control of aquatic macrophytes. Although the targeted plants are generally introduced species, the grass carp is not highly selective and often prefers native plants. It is so voracious that it can completely eradicate nontarget species from particular water bodies before foraging on the target species. Herbivory by introduced European rabbits (*Oryctolagus cuniculus*) has caused enormous damage on islands worldwide. Their impact in Australia is legendary. The most serious damage there is by the stripping and killing of seedling trees and perennial shrubs. In addition, rabbits often cause extensive erosion.

Innumerable crop plants, most of them introduced, have been devastated by introduced insects. Damage in the United States from the Russian wheat aphid (*Diuraphis noxia*) alone exceeded $600 million between 1987 and 1989. The South American cassava mealybug (*Phenacoccus manihoti*) has invaded most cassava-growing regions of Africa and causes yield losses of up to 84%. Similarly, introduced insects attack both native and introduced trees, often with staggering impact. Losses in eastern U.S. forests to the European gypsy moth (*Lymantria dispar*) were estimated at $764 million in one year alone. The Asian balsam woolly adelgid (*Adelges piceae*) has eliminated the Fraser fir (*Abies fraseri*) in many areas of the southern Appalachian Mountains.

Herbivorous insects have been employed in biological control projects against many introduced aquatic and terrestrial weeds. Among striking successes are the use of the South American cactus moth (*Cactoblastis cactorum*) against pest prickly pear (*Opuntia* spp.) in Australia and the South American alligatorweed flea beetle (*Agasicles hygrophila*) on aquatic alligatorweed (*Alternanthera philoxeroides*) in Florida. In each instance, the weed had covered vast areas before the insect introduction, which has subsequently limited it to small, ephemeral infestations. Such introductions of insects for control of introduced weeds have occasionally led to threats to the very existence of nontarget native species. For example, the cactus moth, introduced to control pest prickly pear on the island of Nevis in the West Indies, has reached the Florida Keys, where it attacked every individual of the native semaphore cactus (*Opuntia spinosissima*). The Eurasian weevil *Rhinocyllus conicus*, introduced to North America to control Eurasian pest thistles, especially musk thistle (*Carduus nutans*), attacks several native thistles, including the endangered Suisun thistle (*Cirsium hydrophilum* var. *hydrophilum*). The key to these unplanned threats to native species, in each instance, is that the introduced insect could maintain high numbers on alternative hosts (generally the target species), so that decline of the native species did not induce a decline in populations of the herbivore.

E. Parasitism and Disease

Introduced pathogens of various sorts have had enormous impacts. The most damaging to entire ecosystems are plant pathogens that affect dominant native plants; in so doing, they affect entire communities that interact with these plants. The Asian chestnut blight fungus (*Cryphonectria parasitica*) reached New York City in nursery stock in the late nineteenth century. In less than 50 years, it spread over about 100 million ha of the eastern United States, destroying the aerial parts of almost all mature chestnut (*Castanea dentata*) trees. Because chestnuts had been the most common tree in many forests, the ecosystem impacts of this invasion were almost certainly enormous, although few crucial data were gathered. Several insect species that are host-specific to chestnut went extinct, and nutrient cycling rates were probably greatly modified.

A pathogen of animals that affected many entire ecosystems was the virus rinderpest, native to India. Introduced to Africa in cattle in the 1890s, it infected many native ungulate species, with mortality in certain species reaching 90%; the distribution of some species remains affected a century later. Because of the crucial role played by ungulates in aspects of vegetation structure and dynamics, the impact of rinderpest far surpassed the death of individuals it infected.

Many introduced diseases have heavily affected individual species or small groups of them without apparent major impact on entire ecosystems. Ranges of Hawaiian native birds have been drastically circumscribed by habitat destruction, but avian malaria, caused by *Plasmodium relictum capristranoae* and vectored by introduced mosquitoes, afflicts the remaining populations and helps restrict them to upper elevations. The plasmodium was introduced with Asian songbirds. The European fish parasite *Myxosoma cerebralis* causes whirling disease in salmonid fishes. Following World War II, live North American rainbow trout (*Salmo gairdneri*) were transferred freely among European sites. There they acquired the parasite from the native brown trout, *S. trutta,* which has significant resistance. Eventually, frozen European rainbow trout were widely exported. The parasite probably reached North America in imported frozen fishes. Either these were fed accidentally to fish in a trout hatchery in Pennsylvania, or the viscera may have been discarded in streams near the hatchery. In any event, hatchery fishes became infected, and these were shipped to many other states. In large parts of Montana and Colorado, the great majority of rainbow trout contract the disease, and sport fisheries have collapsed in both states.

Introduced macroparasites also often have major impacts. The parasitic plant witchweed (*Striga asiatica*) probably invaded North Carolina from Africa with military equipment after World War II. It attacks primarily grasses, including corn, and is a sufficient agricultural pest that it has been the target of a long eradication campaign and associated quarantine. The trematode *Cyathocotyle bushiensis*, which causes heavy mortality in ducks, has advanced along the St. Lawrence River recently, concurrently with its invasive introduced intermediate host, the Eurasian faucet snail *Bithynia tentaculata*.

Many parasitic wasps and flies, as well as some disease pathogens, have been introduced as biological control agents for introduced pests. For example, the yellow clover aphid (*Therioaphis trifolii*) is controlled in California by three parasitic wasps, *Praon palitans, Trioxys utilis*, and *Aphelinus semiflavus*. Populations of the cassava mealybug in Africa, discussed earlier, have been greatly reduced by the introduced South American parasitic wasp *Epidinocarsis lopezi*. Perhaps the best-known biological control pathogen is the New World myxoma virus, introduced to mainland Europe (where the European rabbit is native), Great Britain, and Australia (where it is introduced). Initial epizootics caused over 99% mortality in Great Britain and Australia, and over 90% in France. The initial virulent strains in all three countries largely evolved to more benign strains and, at least in Great Britain and Australia, rabbits evolved a degree of resistance. Thus, successive epizootics caused decreasing mortality, until an equilibrium appeared to have been reached. A number of pathogenic fungi have been introduced to control weeds, with varying degrees of success.

F. Hybridization and Introgression

Introduced species can eliminate native species by mating with them, a particularly strong threat when the native species is not as numerous as the introduced one. Both the New Zealand grey duck (*Anas superciliosa superciliosa*) and the Hawaiian duck (*A. wyvilliana*) are thus threatened by extensive hybridization and introgression with the North American mallard (*A. platyrhynchos*), introduced as a game bird. Similarly, the white-headed duck (*Oxyura leucocephala*), now restricted in Europe to Spain, is threatened there by hybridization and introgression with North American ruddy ducks (*O. jamaicensis*), which were introduced to Great Britain as an amenity, escaped, and eventually reached Spain.

Both plants and animals are threatened by such introgression, and its extent is just becoming known with the extensive use of molecular techniques that can detect it. This problem is much more common in regions that exchange closely related species, such as Europe and

North America, than in those with species so distantly related that they are unlikely to be able to mate and exchange genes, such as Australia and either Europe or North America. Exchange of genes is not even necessary for hybridization with an introduced species to affect a native species inimically. Many females of the endangered European mink (*Mustela lutreola*) hybridize with male introduced American mink (*M. vison*), which become sexually mature earlier than the male European mink. The embryos are all aborted, but the loss of reproduction by the European mink exacerbates their population decline.

IV. INDIRECT EFFECTS

The foregoing effects of introduced species are direct effects of various sorts. The actions of individuals of the introduced species are directly on individuals (and often, ultimately, on populations) of one or more native species; they may attack them, eat them, poison them, infect them, and so on. However, an introduced species can also affect other species indirectly in many ways.

A. Classic Indirect Effects

A strictly defined indirect effect occurs when one species alters the interaction between two others. For example, the chestnut blight led to the replacement of chestnut in much of the eastern United States by oak species. Red oak (*Quercus rubra*) increased greatly, and this species is particularly susceptible to oak wilt disease (*Ceratocystis fagacearum*). The increase in oak wilt disease, in turn, raised the frequency of oak wilt disease on many less susceptible native oak species. Thus, the chestnut blight fungus indirectly affected the interaction of oaks and oak wilt disease, in addition to its direct effect on chestnut.

Classic indirect effects can be highly complex. For example, the mite *Pyemotes ventricosus*, accidentally introduced to Fiji, attacked the larvae and pupae, but not the eggs and adults, of the coconut leaf-mining beetle (*Promecotheca coeruleipennis*). Adult beetles then oviposited and died, so the beetle population came to have synchronous, nonoverlapping generations. The absence of larvae and pupae for extended periods caused the mite population to plummet, as did those of two native parasitoids that had previously controlled the beetle; they did not live long enough to survive the intervals between occurrences of the host stages needed for oviposition. So the mite indirectly affected the interaction between the beetle and the parasitoids.

B. Chain Reactions

A number of the introductions discussed earlier affected one species directly, but this direct effect generated subsequent impacts in chain-like fashion. Thus, the direct effect of chestnut blight on American chestnut led indirectly to the extinction of several insect species that were host-specific on chestnut. The modification of nutrient cycles as a chestnut-dominated litter was replaced by litter dominated by leaves of oaks and other species almost certainly affected populations of litter inhabitants as well as plants, but these effects were never investigated. Rabbits in Australia directly affected certain plant species, as noted earlier, and also probably directly contributed to the elimination of two small burrowing marsupials, the boodie rat (*Bettongia lesueur*) and the bilby (*Macrotis lagotis*) by competition for burrows. However, rabbits indirectly led to the elimination of the common wombat (*Vombatus ursinus*) from part of its range by modifying succession and locally eliminating native perennial plants. On coastal islands, the erosion and vegetation changes caused by rabbits have been highly detrimental to seabirds that use these islands for resting and breeding.

Some chain reactions induced by introduced species are so complicated that predicting them would have been difficult. Caterpillars of the large blue butterfly (*Maculina arion*) in Great Britain required development in underground nests of the ant *Myrmica sabuleti*. The ant cannot nest in overgrown areas. Changing land use patterns and reduced livestock grazing left introduced rabbits as the main grazer maintaining the habitat. The biological control introduction of myxoma virus reduced rabbit populations sufficiently so that ant populations declined, and the butterfly went extinct.

Chain reactions can generate effects far from the site of the initial introduction. For example, landlocked kokanee salmon (*Oncorhyncus nerka*) were introduced to Flathead Lake, in western Montana, in 1916 and largely replaced native cutthroat trout (*O. clarki*) as the dominant sport fish. By 1931, the kokanee were spawning in McDonald Creek (Glacier National Park), some 100 km upstream from Flathead Lake, and the spawning population was so large that it soon attracted large populations of bald eagles, grizzly bears, and numerous other predators. The catch of kokanee rose to over 100,000 fishes per year through 1985. Between 1968 and 1975, the opossum shrimp (*Mysis relicta*), native to several large, deep lakes in North America and Sweden, was introduced to three lakes in the upper portion of the Flathead catchment in a misguided attempt to enhance kokanee productivity. The shrimp drifted downstream and reached Flathead Lake by 1981. There was an immediate, dramatic decline in the densities of copepods and especially cladocerans preyed on by the shrimp. This competition with the kokanee for prey caused the kokanee population to decline rapidly; there was no catch at all in 1988 and 1989. Bald eagle numbers fell precipitously, as did those of grizzly bears and several other predators.

C. Vectoring of Pathogens

In a number of instances discussed earlier, an introduced species carried a pathogen that greatly affected one or more native species—rinderpest, whirling disease, avian malaria. In some cases, these could be classified as classic indirect effects. For instance, in Hawaii the avian malaria plasmodium changed the interaction between introduced mosquitoes and native forest birds to the detriment of the latter. The impact of an introduced pathogen can be more complex. For example, the introduced species that carries it to a new region need never be sympatric with a native species ultimately affected. Consider the introduction of the grass carp to the United States. The carp was introduced to Arkansas in 1968; it spread to the Mississippi River, carrying with it a parasitic tapeworm from Asia, *Bothriocephalus acheilognathi*. The tapeworm quickly infested other fishes, including the red shiner (*Notropis lutrensis*), a popular bait fish. Fishermen or bait dealers introduced infected red shiners to the Colorado River, and by 1984 they had reached the Virgin River, a Utah tributary. There they infected the woundfin (*Plagopterus argentissimus*), a native minnow already threatened by dams and water diversions. Tapeworm infections caused woundfin numbers to decline precipitously, possibly because they are less able to compete with the red shiner, which have some resistance to the tapeworm.

V. INVASIONAL MELTDOWN

Some ecological theories suggest that introduced species should interfere with one another and thereby lessen one another's impact. The theory of island biogeography, for example, implies that each successive species in a series introduced to an island (or a habitat island) has a lower probability of surviving. Biological control practitioners have argued about whether releasing several species of potential natural enemies of a single pest might lessen the overall impact on that pest because of competition among them. However, positive interactions between introduced species are detected as often as negative ones. Sometimes one introduced species facilitates the existence and enhances the effect of another; other times the combined effects of two or more introduced species exceed the sum of what the same species might have accomplished individually. Such situations, in which different introduced species enhance one another's effects, are collectively termed "invasional meltdown." The existence of many such cases suggests that an accelerating wave of invasions with ever-increasing effects may characterize the future in many regions.

A. Introduced Animals Pollinating and Dispersing Introduced Plants

Many introduced fig (*Ficus*) species are frequently planted as ornamentals in south Florida. If the wasp that pollinates a fig species in its native range is absent, the fig cannot reproduce. Thus, until recently, figs in Florida remained where planted. However, within the past 15 years, breeding populations of three host-specific pollinating wasps have been introduced to Florida, and the species they pollinate now regularly produce seeds. *Ficus microcarpa* is spreading most quickly and has become an invasive weed, the small fruits of which are dispersed by birds and ants. The spread of the same complex of *F. microcarpa* trees and its wasp pollinator *Parapristina verticillata* is also occurring in Bermuda, Mexico, and Central America.

Some introduced animals disperse introduced plants that disrupt native plant communities. The red-whiskered bulbul (*Pycnonotus jocosus*) has dispersed a number of alien species, such as *Rubus alceifolius, Cordia interrupta*, and *Ligustrum robustum* on the island of La Réunion. *Cordia* was not viewed as a problem until it was widely distributed by the bulbul; the bird is the primary dispersal agent for *Ligustrum*. In the Hawaiian Islands, introduced pigs selectively eat and thus disperse several invasive introduced plant species, and their rooting and defecation favor several introduced invertebrates. Further, the pigs grow larger because of the presence of introduced, protein-rich European earthworms in the soil. The Asian myna bird (*Acridotheres tristis*), introduced to control pasture insects, has dispersed the New World weed *Lantana camara* widely in the lowlands of the Hawaiian Islands, including into native forest areas.

B. Introduced Plants That Modify Habitat

As noted in a previous section, the Atlantic nitrogen-fixing shrub *Myrica faya* may facilitate invasion by other introduced plants that are currently limited by the nitrogen-poor volcanic soil and absence of native nitrogen-fixers. This plant also enhances populations of introduced earthworms, which in turn increase the rate of nitrogen burial and thus exacerbate the impact of the plant on nitrogen cycling. The soil drying by introduced salt cedars already described favors the introduction of nonnative grasses in both the U.S. Southwest and Australia. The increased soil salinity produced by the African crystalline ice plant, described previously, does more than kill native plants. When wind or other disturbances create holes in the carpet of ice plant, they are colonized not by native plants but by the ice plant itself or by weedy introduced plants such as *Malva parviflora* or *Erodium cicutarium*.

Many invasive introduced plants increase fire frequency and/or intensity to the detriment of native species but to their own advantage and that of other invaders. Old World grasses have come to dominate many New World grasslands through this facilitation. For instance, in Hawaii the introduced perennial grass *Schizachyrium condensatum* invaded seasonal submontane shrub-dominated woodland. It fostered much more frequent fires over larger areas, killing most native trees and shrubs. But *S. condensatum* recovered quickly and the even more flammable introduced perennial grass *Melinis minutiflora* also invaded.

C. Introduced Animals That Modify Habitat

The disturbance wrought by large, congregating, introduced herbivores in North America favored the establishment of Eurasian grasses adapted to such animals, and helped them to replace native tussock grasses. Similarly, the Asian water buffalo (*Bubalus bubalis*), introduced to northeastern Australia as a beast of burden and for meat, spread throughout the flood-plain of the Adelaide River by the late nineteenth century. It devastated native plant communities, compacted the soil, eroded creek banks, and altered hydrology. A Central American shrub, *Mimosa pigra*, had been a minor, noninvasive weed in the vicinity of Darwin for a century. This legume produces large numbers of small seeds that are readily dispersed by water. The sundering of the flood-plains by the water buffalo created ideal germination habitat for *M. pigra* seedlings. In many areas, native sedgelands have been converted to a monoculture of *M. pigra*.

The Caspian zebra mussel (*Dreissena polymorpha*) arrived in the Great Lakes in the 1980s. Its huge numbers and great filtering ability have greatly affected native freshwater communities over much of the eastern and midwestern United States. In particular, they convert large amounts of seston into excreted feces, creating a soft substrate of rich organic material. Among species whose populations have increased as a result of the zebra mussel is the invasive Eurasian faucet snail discussed previously. The filtering also increases water clarity, and this change has favored certain invasive macrophytes, such as Eurasian watermilfoil (*Myriophyllum spicatum*).

D. Mutualisms

A number of instances in which one introduced species facilitates the existence, spread, and population growth of another introduced species constitute mutualisms, in that the latter species also aids the former. For example, introduced macrophytes like Eurasian watermilfoil provide additional settling substrate for the mussel, and they can also help the mussel disperse between water bodies. *Mimosa pigra* aided water buffalo by protecting them from aerial hunters trying to eradicate them. Fig wasps and their associated fig species obviously form coevolved mutualistic introduced species pairs. In the Hawaiian Islands, introduced African big-headed ants (*Pheidole megacephala*) tend an introduced scale insect, *Coccus viridis*; this is a classic mutualism. The scale is on the introduced plant *Pluchea indica*; among other things, the ant hinders introduced predatory coccinellid beetles and parasitic wasps. In this one case, a range of interactions are taking place among an entire assemblage of introduced species. Sometimes, as with *M. pigra* and water buffalo, the mutualists cannot have coevolved, as they originate on different continents. The big-headed ant in Hawaii also tends the South American gray pineapple mealybug (*Dysmicoccus neobrevipes*), a particularly unwelcome pest because it helps spread a wilt disease of pineapple.

VI. TIME LAGS AND EVOLUTION

Effects of introduced species may change dramatically, for example, when an invader is quiescent during a time lag, after which it rather abruptly spreads and increases in number. As observed in Section I, such time lags complicate efforts to predict the effects of an introduced species and are sometimes mysterious. It is frequently suggested that a mutation in an invader could account for a particular time lag, and that evolution in general could greatly change the effects of an introduced species. In no instance can a sudden increase in population size and impact of an introduced species be clearly attributed to a mutation, but there is ample evidence that such species evolve in their new homes. How frequently and to what extent does such evolution change their effects?

Morphological evolution of introduced species is often apparent. The small Indian mongoose, whose depredations of native island species were discussed earlier, has become larger and more sexually dimorphic on all islands to which it has been introduced. This change is probably selected for in the absence on the islands of the slightly larger gray mongoose (*Herpestes edwardsii*), with which the small Indian mongoose is sympatric in the region of origin for the island populations. The North American muskrat (*Ondatra zibethicus*), introduced as a fur bearer to Europe in 1905, has spread throughout much of Europe in less than a century. The muskrat is well known in its native range as a species with substantial variation in morphology (especially size), to the extent that various subspecies have been named. In Europe, it has already evolved approximately the same degree of variation. Similarly, the European house sparrow (*Passer domesticus*) was introduced in New York City in 1853 and quickly spread to become one of the most common birds in North America, covering much of the continent. It has evolved so much that distinct "races" are now easily identified in different parts of its introduced range. Many plant species are known to differ substantially between their native and introduced ranges, though there is generally far less evidence than in the vertebrate cases just cited that the differences are genetic. What is rarely if ever known in these demonstrated instances of evolution after introduction is how these changes affect the impacts of these species on native communities and ecosystems.

The hybridization of an introduced species with a native one can even lead to the formation of a new species, which can be invasive. For example, North American cordgrass (*Spartina alterniflora*), introduced in shipping ballast to southern England, occasionally hybridized with a noninvasive native congener, *S. maritima*. These hybrids were sterile, but eventually one hybrid individual underwent a doubling of chromosome number to produce a fertile new species, *S. anglica*, which turned out to be highly invasive.

The evolution of the introduced rabbit and myxoma virus in Australia and Great Britain certainly affected the impact of the rabbit on native systems. The fact that the rabbit became somewhat more resistant to the virus, and that the virus evolved to be somewhat more benign, lowered the rabbit mortality in each successive epizootic, thus the degree of various effects outlined earlier. As pathogens and their hosts often coevolve, changes of impact might be expected. They are not always seen, however. Chestnut blight has not become perceptibly less devastating to American chestnut, nor avian malaria to native Hawaiian birds.

Many introduced pest insects have evolved resistance to chemical pesticides and thus generate much greater impacts, generally on agricultural or silvicultural plants. Just as strains of human pathogens have evolved resistance to whole sequences of antibiotics, some pest insects have evolved resistance to several insecticides. This resistance can arise in three ways: insects can evolve to tolerate greater amounts of the chemical, they can evolve physiological means of detoxifying the chemical, or they can evolve behavioral traits (such as going to the bottom of a leaf instead of the top) that help to avoid contact with the chemical. Introduced weedy plants also evolve resistance to herbicides, though this phenomenon has not been as widely studied as insect resistance.

The evolution of resistance to chemicals by introduced pests gives impetus to the desire to control them biologically, through the introduction of natural enemies. However, this approach has fostered concerns about subsequent evolution. For pathogens, one concern is that the target organisms will evolve resistance to them. Over 90% of the market for insect pathogens today consists of products involving two organisms, the bacterium *Bacillus thuringiensis* and heterorhabditid nematodes. Both types have no natural association with insect pests on plants and thus have to be routinely reapplied. It is possible that target pest insects could become resistant to them simply through repeated exposure plus selection of resistant strains. This possibility is heightened for the bacterium by the fact that *B. thuringiensis* toxins are now being genetically engineered into crop plants, thus increasing the exposure of pest insects to them and increasing the rate of natural selection. On the other hand, introduced biocontrol insects can evolve greater adaptation to a potential host and thus become more rather than less lethal. In the United States, the Egyptian alfalfa weevil (*Hypera brunneipennis*) was originally quite immune to the European ichneumonid parasitic wasp *Bathyplectes curculionis*, encapsulating 35–40% of the eggs and larvae in an immune response. Fifteen years later, only 5% of the eggs were encapsulated.

A second concern is that introduced biological control agents could evolve to switch hosts. For introduced insects, the fact that host range can be controlled by a single gene only heightens this concern, as does the fact that certain features of most biological control introductions favor fast evolution. These features include initially small population size, rapid population growth in new environments, and novel, abundant resources. In light of the possibility, it might seem surprising that few examples of host switching are known. Because many of the pathogens used or contemplated as biological control agents have broader host ranges than parasitic insects, the probability of host-switching accompanied or followed by evolutionary adaptation to the new host is enhanced.

VII. QUANTIFYING EFFECTS

"Effect" and "impact" have never been formally defined in invasion biology, so that rankings of introduced species in terms of greatest impact are impressionistic. Area of occupancy is a frequently used index of effect, although the number of individual invaders per unit area and what each individual actually does would be important (though more difficult to measure). Impacts on ecosystem processes, such as nutrient cycling and fire frequency, are seen by some biologists as the best measures of effect of introduced species, on the grounds that such alteration of processes could affect a large number of species simultaneously. Both areal coverage and impact on processes often envision outcome for native species, communities, and ecosystems as the crucial measure of effect. The real effect, in these terms, would require accurate knowledge of the range size of the species, the distribution of abundance of species within this range, and the impact on the native species, both on-site and elsewhere, per individual invader (or per unit biomass of the invader). The first two factors can be measured or at least estimated in straightforward fashion, although the effort for some species would be enormous. The per capita impact, on the other hand, would probably entail not only detailed observation but also experiments, perhaps over the long term, at individual, population, community, and ecosystem levels. As one example of the subtleties that might be involved, the fact that an introduced predator is observed to eat individuals of a native species does not necessarily mean that this predation affects the population size, behavior, or ecology of the prey species. Ecological methods are well enough developed that these sorts of questions can be answered, but they generally require detailed research, not quick observation.

Economic costs of an invasion are often proposed as estimates of the effect of introduced species, but these costs are more easily tallied from a human standpoint than from that of native species, communities, and ecosystems. It is a fairly straightforward matter to tabulate the costs expended in controlling a particular weed or insect pest, in terms of materiel and personnel used. And one could relatively easily add the monetary cost of many services or goods lost because of an introduced species—say, a percentage of some crop. Costs of other lost services are not so easily tabulated, even in human terms. For example, how expensive is the loss in enjoyment suffered by citizens who can no longer walk through Fraser fir or American chestnut forests in eastern forests because these were eliminated by introduced species? Economists have attempted to measure these costs. For example, travel costs of citizens to a particular natural site, as evidence of how much they are willing to pay for such an experience, are sometimes used to estimate the "value" of a site, at least to humans. But all of these methods are controversial. Furthermore, costs to nonhuman species are even harder to weigh in monetary terms. What was the cost, in dollars, to the insect species extinguished because of the chestnut blight?

In sum, even though a bewildering variety of effects of introduced species can be demonstrated, measuring and comparing these effects for different species remains as much a conceptual and practical challenge as predicting the effects.

See Also the Following Articles

COEVOLUTION • FIRES, ECOLOGICAL EFFECTS OF • GRAZING, EFFECTS OF • HUMAN IMPACT ON BIODIVERSITY, OVERVIEW • INTRODUCED PLANTS, NEGATIVE EFFECTS OF • ISLAND BIOGEOGRAPHY • SPECIES COEXISTENCE

Bibliography

Biological Invasions (journal. Volume 1, 1999).

Bright, C. (1998). *Life Out of Bounds. Bioinvasions in a Borderless World.* W. W. Norton. New York.

Carey, J. R., Moyle, P., Rejmanek, M., and Vermeij, G. (eds.). (1996). Invasion Biology (special issue). *Biol. Conservation* 78, 1–214.

Cronk, Q. C. B., and Fuller, J. (1995). *Plant Invaders.* Chapman & Hall, London.

Drake, J. A., Mooney, H. A., di Castri, F., Groves, R. H., Kruger, F. J., Rejmanek, M., and Williamson, M. (eds.) (1989). *Biological Invasions. A Global Perspective.* John Wiley & Sons, Chichester, United Kingdom.

Elton, C. S. (1958). *The Ecology of Invasions by Animals and Plants.*

Methuen, London. Reprinted 2000 by University of Chicago Press, Chicago.

Kareiva, P. (ed.). (1996). Special Feature. Developing a predictive ecology for non-indigenous species and ecological invasions. *Ecology* 77, 1561–1697.

Knight, B. N. (ed.). (1993). *Biological Invasions. The Control and Impact of Invasive Exotic Species.* Indiana Academy of Science, Indianapolis.

Sandlund, O. T., Schei, P. J., and Viken, A. (eds.). (1996). *Proceedings of the Norway/UN Conference on Alien Species.* Directorate for Nature Management/Norwegian Institute for Nature Research, Trondheim, Norway.

Simberloff, D., Schmitz, D. C., and Brown, T. C. (eds.). (1997). *Strangers in Paradise. Impact and Management of Nonindigenous Species in Florida.* Island Press, Washington, DC.

U.S. Congress, Office of Technology Assessment. (1993). *Harmful Non-indigenous Species in the United States.* U.S. Government Printing Office, Washington, DC.

Williamson, M. (1996). *Biological Invasions.* Chapman & Hall, London.

INVERTEBRATES, FRESHWATER, OVERVIEW

Margaret A. Palmer* and **P. Sam Lake**†
**University of Maryland and †Monash University, Australia*

GLOSSARY

benthic At or on the bottom of streams, lakes, or riverbeds.

decomposition The breakdown of organic matter such as dead plants and algae to release carbon and nutrients.

endemic species A species that is unique to a particular locality.

exotic species A species that has been introduced to an area outside of its native range.

groundwater A "reservoir" of water residing below ground in saturated soils and beneath geologic formations.

larvae Early stages in the development of an organism; for invertebrates, often morphologically quite distinct from the adult.

local species richness Number of species found at a local site; distinguished from regional or global species richness that "sums" the number of species across a number of individual sites.

planktonic Organisms that reside in the water column.

watershed A geographical region in which water drains into common water bodies.

WATER IS THE MOST abundant substance on earth and is essential for all of life. It is critical to the world's climate, to the cycling of nutrients, and is a habitat for much of the earth's biodiversity. Fresh water links the land and oceans via groundwater and riverine flow. While invisible to the human eye, groundwaters represent a substantial portion of the fresh waters (approx. 30%) acting not only as the most important reservoir of fresh water on earth, but as a home to many unique fauna. Lakes and reservoirs are also important stores of fresh water and harbor several groups of extremely abundant invertebrates. Running waters and wetlands are known to be havens for diverse assemblages of invertebrates many of whom are important in ecological processes that ensure clean water. The amount and timing of water inputs to many freshwater systems worldwide is changing rapidly. Since many invertebrates require specific habitats and flood regimes, their abundance and diversity in many parts of the world are at risk because habitats are being lost and natural flow regimes altered at astounding rates. Also contributing to the recent declines in freshwater invertebrates are poor water quality and the introduction of exotic species that have led to the extinction of native species, especially bivalves and crayfish.

I. INTRODUCTION

A. Major Freshwater Habitats

The major freshwater habitats include running waters (streams and rivers), standing or semistanding waters

Encyclopedia of Biodiversity, Volume 3

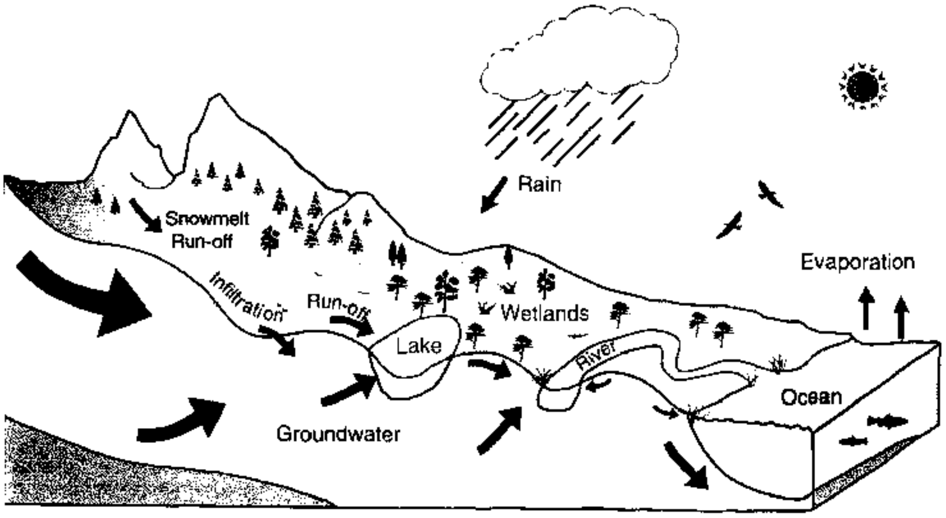


FIGURE 1 Schematic illustration of major freshwater habitats. Note all water bodies, including lakes, streams and rivers, wetlands, and the oceans, are connected via the groundwaters. See also color insert, Volume 1.

(reservoirs, lakes, ponds, wetlands), and groundwater (Fig. 1). Most of the 35 million cubic kilometers of fresh water on the earth (≈ 69%) exists in the form of polar ice. Of the remaining fresh water, a huge volume exists out of sight as groundwater (Table I). Lakes, streams, and wetlands represent less total water volume but are biologically rich hot spots for abundant and diverse assemblages of invertebrates. Together, the freshwater bodies form a continuous "pipeline" that links the atmosphere, land, and the oceans. Rainwater is intercepted by vegetation on land, percolates through the soils into groundwaters, and then moves into wetlands, lakes, and streams before it eventually reaches the oceans (Fig. 1). Since groundwater directly or indirectly "feeds" all other water bodies on earth, maintenance of adequate quantities of high-quality water in this underground reservoir is a high priority worldwide.

B. Major Groups and Their Lifestyles

A disproportionate number of species live in fresh waters compared to marine systems: oceans comprise 70% of the earth's surface while inland waters make up about 1%. Despite this, about 15% of all animal species alive today live in fresh water. Most of the freshwater animal species are invertebrates. More than 70,000 species have been described but many more remain undiscovered—new species and genera are found every year. Freshwater habitats have representatives from most taxonomic groups with the insects being particularly speciose (Table II). The freshwater insects are also quite notable for

TABLE I

Freshwater Distribution by Continent

	Africa	Europe	Asia	Australia	North America	South America
Lakes and reservoirs	31,240	2,447	29,132	192	26,573	1,199
Rivers and streams	195	80	565	25	250	1,000
Groundwater	5,500,000	1,600,000	7,800,000	1,200,000	4,300,000	3,000,000
Wetlands	341,000	Eurasia	925,000	4,000	180,000	1,232,000

Modified from the World Conservation Monitering Center's *Freshwater Biodiversity: A preliminary global assessment* (1998). Data are volume of water in km^3; wetlands includes marshes, swamps, lagoons, and floodplains.

TABLE II

A List of Some of the Most Common Freshwater Invertebrates with Estimates of the Number of Described Species Globally

	Common name	Aquatic stage	Existence	Global diversity
Phylum Porifera	Sponges	SW, RW	B	150
Phylum Cnidaria	Hydra, jellyfish	SW, (RW)	B, P	30
Phylum Platyhelminthes				
Class Turbellaria	Flatworms, Planaria	RW, SW, GW	B	1000
Phylum Annelida				
Class Hirudinea	Leeches	RW, SW	B	350
Class Oligochaeta	Earthworms	RW, SW, GW	B	700
Phylum Nematoda	Free-living roundworms	RW, SW, GW	B	1500
Phylum Gastrotricha		RW, SW	B	200
Phylum Rotifera	Wheel-animalcules	RW, SW	B, P	2000
Phylum Tardigrada	Water bears	SW, (RW)	B	
Phylum Arthropoda				
Class Crustacea				
O. Cladocera	Water fleas	SW, (RW)	P, B	400
O. Ostracoda	Seed shrimp	SW, GW, (RW)	B	1000
O. Copepoda		SW, GW, RW	B, P	2000
O. Isopoda	Sow bugs	SW, RW, GW	B	1000
O. Amphipoda	Scuds	SW, GW, RW	B, P	1500
O. Decapoda	Crabs, shrimp, crayfish	SW, RW, GW	B, P	1300
Class Arachnida				
Subclass Acari	Mites	SW, GW, RW	B	5000
Class Insecta				
O. Ephemeroptera	Mayfly larvae	RW (SW).	B	2200
O. Odonata	Dragonfly and damselfly larvae	RW, SW	B	4900
O. Plecoptera	Stonefly larvae	RW, (SW)	B	2100
O. Hemiptera	true bugs	RW, SW (GW)	B	3200
O. Trichoptera	Caddisfly larvae	RW, SW	B	7000
O. Lepidoptera	Butterfly and moth larvae (caterpillars)	SW, RW	B	100
O. Coleoptera	Beetles	SW, RW	B	5000
O. Megaloptera	Dobsonfly larvae	SW, RW	B	300
O. Neuroptera	Spongillaflies	On sponges	B	100
O. Diptera	Fly larvae (e.g., mosquitos, midges)	SW, RW, GW	B, (P)	>20000
Phylum Mollusca				
Class Gastropoda	Snails	RW, SW, GW	B	2500
Class Bivalvia	Clams, mussels	RW, SW	B	1400

Some groups are more common in running waters (RW), such as streams and rivers while others are more common in standing or semistanding water (SW) bodies (lakes, streams, wetlands) or groundwater (GW). Some are primarily benthic (B), existing in or on the bottom of lake, stream, river sediments, or on aquatic vegetation, while others maintain a water column existence as members of the plankton or pelagic fauna (P). Parentheses are used when an organism may be found in a habitat type but is less common there than in the other habitats.

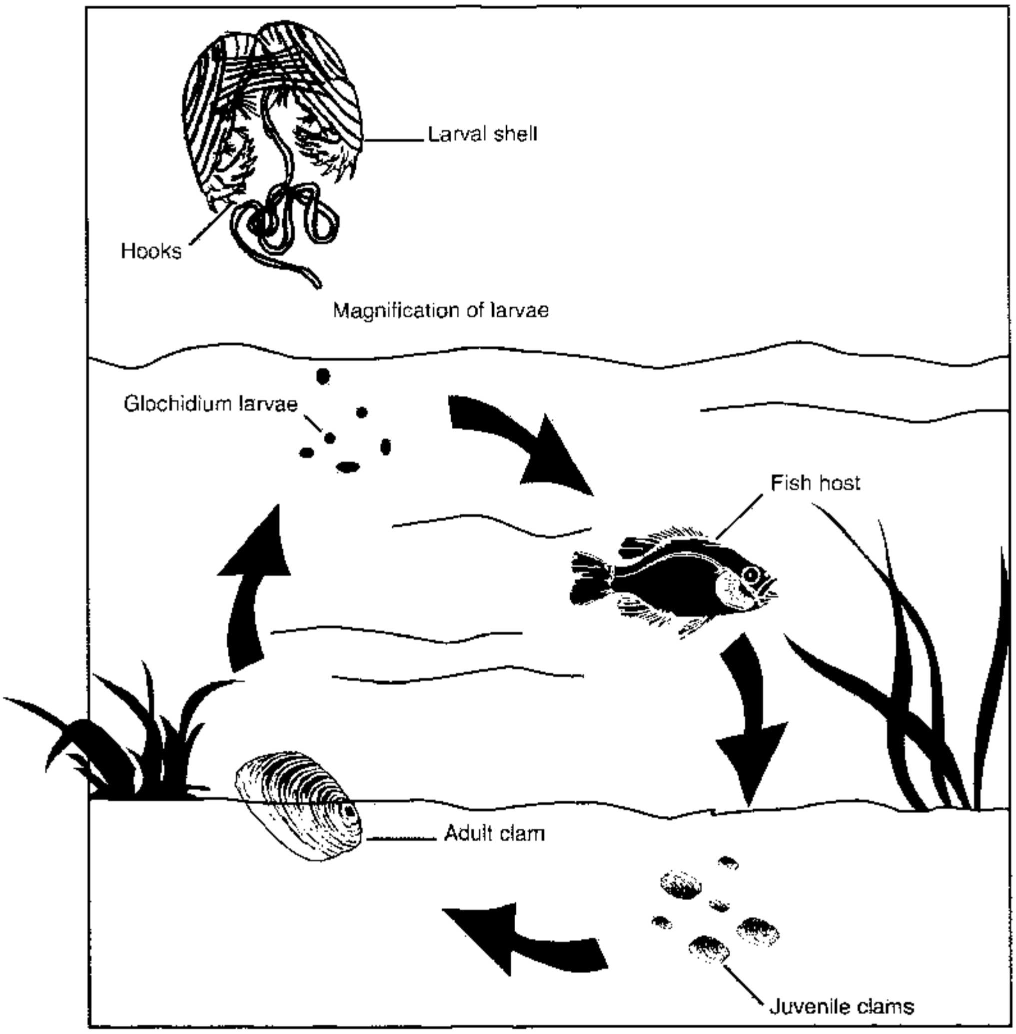


FIGURE 2 Example of a complex life cycle for a freshwater invertebrate. Many freshwater bivalves are benthic yet produce larvae that are released into the water column. These develop into small glochidium larvae that attach to a fish host. After some time, the larvae metamorphose into small bivalves, detach, and take on a benthic existence.

the large number whose lifestyle changes with age: many have aquatic larval stages but flying adults. This is in contrast to freshwater mollusks and crustaceans who complete their life cycles in water either as members of the plankton or benthos.

Life cycles of freshwater invertebrates can be quite complex with multiple, morphologically distinct stages (Fig. 2). For example, among the freshwater bivalves, some are hermaphroditic while others have separate sexes; some brood their eggs and then release small clams; however, many release larvae into the water. Zebra mussels produce a planktonic larval stage (a veliger larva), while some of the most common freshwater native bivalves have unique parasitic larvae (glochidia larvae) that attach to the skin, scales, or gills of fish. The host fish disperses the attached glochidium larvae, prior to the detachment of the larvae and their metamorphosis into adult bivalves. This dispersal phase may be very important to the long-term survival of bivalves particularly in freshwater bodies that are being stressed.

Damselflies and dragonflies provide other examples of complex life cycles typical of many freshwater invertebrates. These insects have flying adults that live weeks to months and lay eggs on aquatic vegetation or other substrates. The eggs hatch within 12 to 30 days, or they can remain viable in an unhatched state (diapause) if there is a drought. Small larval stages (sometimes called nymphs) hatch from the eggs and then molt multiple times (10–16) before crawling up vegetation and "emerging" as flying adults. The presence of an intact riparian zone with a diversity of native plants promotes population persistence of insects.

C. Number of Freshwater Invertebrate Species

The number of freshwater invertebrates varies considerably between lakes, running waters, groundwaters, and wetlands. Since the number of species varies vastly from locale to locale as a function of many complex factors, estimates of species richness should be taken as only approximations. Additionally, freshwater invertebrate diversity at any one time may be moderate while accumulated diversity over time may be very high. For example, there may be dramatic declines in invertebrate diversity in many streams during the flood season but high levels of diversity across the entire year. Similarly, temporary wetlands may have much higher total diversity over time than permanent wetlands even though the snapshot diversity in a temporary wetland may be considerably lower.

Many freshwater wetlands are extremely speciose habitats (e.g., up to 2000 invertebrate species at a locale), followed by lakes and streams (80–1400 species of invertebrates typically found at a locale), and then groundwaters (0–150 at a locale). In all these habitats, insects are by far the most speciose group with local species richness levels of 50 to 500 and global richness levels of probably at least 45,000 species. Crustaceans may attain local species richness values of up to 150 and global richness values of 8000 species; water mites up to 75 species locally and 5000 species globally; annelids up to 50 species locally and 1000 species globally; nematodes and rotifers up to 500 species locally and 6000 species globally; and mollusks up to 100 species locally and 4000 globally.

Some freshwater habitats are hot spots of endemism harboring many unique fauna. For example, ancient lakes such as Lake Baikal in Siberia has a rich abundance of endemic fauna, especially for those species that live all of their lives in fresh water and have a poor capacity for dispersal. For one group of crustacean amphipods, the gammarids, this lake has the highest level of endemism in the world (41 genera, of which 38 are endemic). Other lakes that harbor highly endemic invertebrate faunas include the African rift lakes of Lake Tanganyika and Lake Nyasa, Lake Titacaca in Peru, and Lake Lanao in the Phillipines.

D. Dominant Taxonomic Groups across Freshwater Habitats

In running waters, the dominant invertebrate groups vary considerably depending on the type of bottom substrate (e.g., mud, cobble/boulder) and the flow (Fig. 3). In rough-bottom creeks and swift-flowing streams, invertebrates live primarily on or beneath pebbles and boulders on the bottom and include many species of insects, crustaceans, and mollusks. Many of these benthic species ("shredders") are adapted to feed on decomposing plant material that has fallen into the stream. Further downstream in the watershed, the channel widens and more sunlight reaches the bottom so that invertebrate species composition changes. Here we find animals that feed on decomposing plants and animals that feed on algae ("scrappers") growing on the rocks and sides of channels. Even further downstream in the watershed, in deep riverine areas, fine sediments predominate, the neabed water flows more slowly, and planktonic invertebrates become common in the water column. In the soft sediments, burrowing insect larvae, bivalves, and worms are common as long as oxygen is present.

In lakes and ponds, the most common invertebrates include crustaceans, rotifers, insects, and oligochaetes; however, taxonomic composition varies dramatically depending on the size of the lake and the position within the lake (Fig. 4). Benthic invertebrates are particularly abundant and speciose along the lake margins but generally are less numerous in the deeper parts of the lakes where oxygen may be limiting and habitat diversity is much lower. However, a few species that are tolerant of silt and low oxygen such as oligochaetes, midge larvae, and nematodes may be extremely abundant in the deeper part of lakes. In the water column, abundant planktonic invertebrate communities thrive by feeding on phytoplankton, but these invertebrate assemblages typically are dominated by just a few taxonomic groups such as rotifers, copepods, and cladocerans.

In the groundwater, crustaceans, rotifers and nematodes are very common, and many of these are adapted for adept movement among rocks and particles of sand. In some parts of the world, insect larvae are common in the groundwater and may be found great distances away from the nearest surface water. Important crustacean groups in groundwater include copepods and less commonly amphipods, ostracods, isopods, and decapods. Many of these live in caves and sinkholes. Groundwater invertebrates are particularly interesting because of the unique adaptations many of them possess for life in dark, isolated habitats where food may be scarce. For example, the Edwards Aquifer in Texas harbors a unique fauna of 22 species including 10 amphipod species. Many of these species have reduced (or no) eyes, reduced body pigmentation, reduced body size, low metabolic rates, and an enhanced sensitivity to touch.

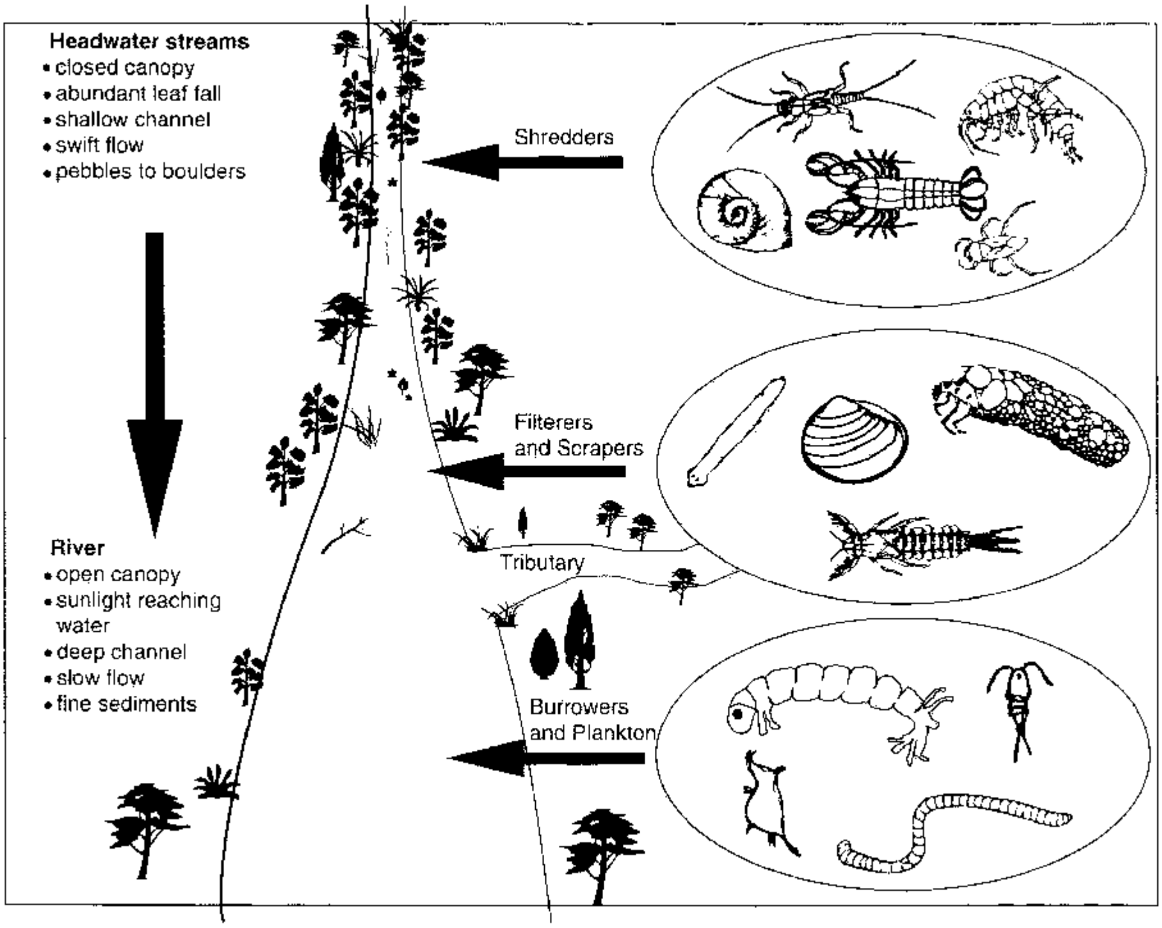
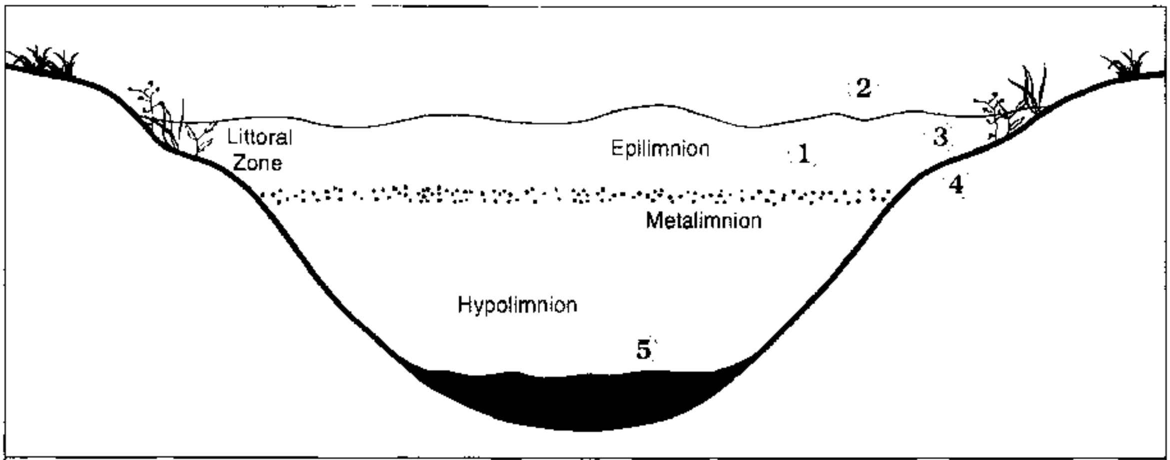


FIGURE 3 Schematic illustrating the relationship between stream size (small upland streams to large downstream rivers), flow and bottom substrate, and type of macroinvertebrate groups likely to dominate. In the smallest streams, large inputs of leaves and wood along with benthic algae are particularly important sources of food for invertebrates. These invertebrates "shred" the leaf material into finer particles and thus facilitate its further breakdown by fungi and bacteria. Further downstream, where sunlight penetrates the water and flow is reduced, algae are also important sources of food and many invertebrates "scrape" the algae from rock surfaces. In the very open riverine areas, plankton may be an important food source and many of the benthic invertebrates burrow into the fine sediments. Modified from Vannote *et al.* (1980).

In freshwater wetlands, the number of habitats and types of food sources for invertebrates are vast. The soft sediments are filled with burrowing nematodes, oligochaetes, midge larvae (a dipteran insect), and mites. Some invertebrates feed on or around the roots of vegetation, while others burrow through the sediments consuming bacteria, fungi, or other invertebrates. In the overlying water, crustaceans, rotifers, and insect larvae (e.g., mosquito larvae) are common and feed on phytoplankton or invertebrate prey. The stems of emergent and submerged vegetation in wetlands are typically crawling with diverse assemblages of crustaceans, mollusks, and insects. Small shrimplike crustaceans (amphipods) may reach very high levels of abundance on aquatic vegetation. Water beetles, Odonates (dragonflies and damselflies), and waterbugs (water boatmen, waterstriders, and backswimmers) are commonly the top predators.

E. Global Patterns in Freshwater Biodiversity

Knowledge of lineage histories is very important because regional patterns of distribution of species are often a reflection of distinctly different evolutionary histories. Adaptations to specific climates and habitat characteristics influence speciation and the ability of an organism that evolved under one set of circumstances to move broadly among different geographic areas. Terrestrial and marine biologists generally have found the highest levels of animal and plant diversity in the tropics with lower levels in temperate and polar regions. This



Some Representative Invertebrates in Lake Habitats

1 Pelagic Zooplankton	2 Surface	3 Littoral Water Column	4 Littoral Benthos	5 Hypolimnion Benthos
protozoa	protozoa	water mites	hydra	gastrotrichs
rotifers	springtails	rotifers	bryozoans	flatworms
cladocera	water spiders	cladocerans	leeches	nematodes
copepods	water mites	water boatmen	true midges	oligochaetes
phantom midges	water striders	phantom midges	ostracods	true midges
			amphipods	

FIGURE 4 Representative invertebrates from different zones within a typical lake: (1) pelagic zooplankton; (2) surface; (3) littoral water column; (4) littoral benthos; (5) hypolimnion benthos. Modified from Thorp and Covich (1991), Figure 2.9.

is not necessarily the case for freshwater invertebrates. Species richness for freshwater invertebrates is related in very complex ways to latitude and elevation and global patterns of diversity are not straightforward for this group.

There have been far fewer surveys of freshwater invertebrates in the tropics than in temperate regions; however, those to date suggest that biodiversity in fresh water is not typically higher in low latitude, tropical habitats. For example, in North America, bivalves and crayfish appear to be most diverse in temperate zones (about 260 native bivalves and 320 species of crayfish). Temperate zones have a wealth of crayfish species, but tropical fresh waters have a richer shrimp and freshwater crab fauna. Insects, mayflies, and stoneflies are more diverse in temperate latitudes than in the tropics, whereas dragonflies, water bugs, and water beetles are much more diverse in the tropics.

II. FACTORS INFLUENCING BIODIVERSITY IN FRESH WATERS

Scientists have devoted lifetimes to the study of what determines the number of species in a given habitat type or geographic region. These studies have involved thoughtful systematic observations of patterns of distribution dating back many years (e.g., Robert Pennak's work in many freshwater systems), and more recent experimental studies that have cleverly manipulated some factor believed to influence biodiversity. From such studies, we know that biodiversity in fresh waters is determined by both historical (e.g., age of a lineage, evolutionary constraints) and contemporary factors (e.g., habitat heterogeneity, predator-prey interactions). The relative importance of the many factors varies geographically as well as between habitats in the same geographic region.

A. The Origin of Freshwater Invertebrates

Most freshwater invertebrates are believed to be derived from terrestrial or marine ancestors. Representatives from many different groups of marine crustacea (shrimp, copepods, cladocera) may have made the transition from salt water to fresh water through many years of evolution in increasingly brackish (low salinity) water. Success in fresh water required that the animals be able to maintain ionic balance in their body fluids (not gain too much body water), reproduce without the aid of complex tidal currents to carry eggs and larvae, and, in some cases, develop life stages that could withstand freezing or drying. Only certain species could make this transition, so at the phylum level, freshwater biodiversity is more limited than in the oceans. For example, echinoderms (e.g., starfish) did not develop the ability to osmoregulate in fresh water.

Many freshwater invertebrates are secondary aquatic forms that moved, in an evolutionary sense, into fresh waters from the land. The most notable group of secondary invaders is the insects, many of which still have terrestrial adults. Some insect orders, such as the Hemiptera, Coleoptera, and Diptera, have both many freshwater species and many terrestrial species. Hemiptera and Diptera also have a few highly specialized marine species. Water mites (Acari) coevolved with insects and many species are parasites on adult insects. With their coevolutionary links, they are also secondary invaders of fresh waters. Freshwater snails consist of two major groups, the prosobranchs and the pulmonates. Pulmonates have invaded freshwater systems from the land; freshwater forms include the two common and ubiquitous families of pond snails, the Lymnaeidae and Physidae.

B. Habitat Age and Isolation

Some freshwater habitats are extremely old and have been isolated for very long periods of time. These habitats have a high degree of endemism. For example, Lake Baikal in Siberia is extremely old and has many species that are found no where else in the world. Such "hot spots" of unique species are considered valuable for conservation purposes and for use as a "natural laboratory" for the study of invertebrate evolution.

For those ancient freshwater bodies that are highly isolated, resident fauna may be at high risk of extinction. If a freshwater species is highly specialized for conditions at a particular site and conditions change, the species may be unable to survive. Geographic or habitat remoteness is not the only factor promoting isolation of freshwater invertebrates. Some fauna are isolated because they have very limited dispersal abilities. For example, many groundwater and deep-streambed dwelling species are very poor swimmers and lack dispersal stages. Thus local biodiversity may be low even though regional biodiversity is much higher—species from the large-scale regional "pool" of invertebrates (e.g., all rivers in a geographic region) may not be able to move between all watersheds. In general, ecologists assume that local species diversity represents a subset of the regional pool of species and that the size of this subset depends greatly on both dispersal abilities of fauna and geographic barriers to dispersal.

C. Habitat Heterogeneity and Disturbance

A central tenet of ecological theory is that the greater the number and types of habitats, the greater the species diversity. Put simply, different types of organisms with different niche requirements can exist in areas the are heterogeneous over space and time. In freshwater ecosystems, spatial heterogeneity (and species diversity) is enhanced by anything that leads to a great variety of particle sizes in which benthic organisms can burrow and feed. Similarly, spatial heterogeneity and species diversity of planktonic invertebrates is enhanced by anything that promotes the growth of many types of vegetation because this provides the invertebrates with diverse food types and hiding places from predators.

Environmental variation over time (temporal heterogeneity) is also extremely important as a determinant of species diversity. Variable climate and flow regimes will promote seasonal (or longer) changes in species abundances and may determine whether a species can coexist in a location. Some freshwater invertebrates have evolved life cycles to take advantage of or avoid predictable changes in water level, flood flows, or ice scour. However, when these environmental changes become extreme or highly unpredictable (a "disturbance"), species richness may decline as fauna are unable to evolve coping strategies that keep up with such changes. Many ecologists believe that invertebrate species diversity is greatest when natural disturbances like floods are at intermediate levels with respect to severity or predictability. High levels of disturbance will wipe out most species while low levels of disturbance may favor one or two species to the exclusion of other species that are poorer competitors.

D. Species Interactions

Predation and competition are well known to influence local species diversity in freshwater ecosystems. For example, many fish predators are highly selective in their feeding and may forage on a single species or suite of species. Thus, predation pressure may reduce species diversity locally, or it actually may promote species diversity if the predator is removing a species that is a competitive dominant. Competitive dominants are species that are so good at garnering resources (space, food, etc.) that other species are unable to coexist with them. The process of competitive exclusion also may be prevented by disturbances if the disturbance (for example, a flood or drought) has a disproportionate effect on the competitive dominant.

In freshwater ecosystems, some of the most dramatic examples of the delicate balance that species interactions play in determining biodiversity come from the introduction of exotic species. Many exotic fish have been introduced into lakes throughout the world and have had devastating effects on their invertebrate prey or on other fish who in turn prey on invertebrates. Introduced carp have reduced invertebrate populations in floodplain wetlands. Non-native trout introduced to streams have depleted populations of native fish and increased trout predation pressure on native invertebrates.

III. THE ROLE OF INVERTEBRATES IN ECOLOGICAL PROCESSES

Healthy freshwater ecosystems are those in which ecological processes continue unimpeded to ensure that water is clean and that organic matter in a lake, stream, or wetland is not lost or accumulated in excess. Freshwater invertebrates play key roles in these processes. The most important ecological processes in freshwater systems include the decomposition of organic matter; the uptake and transfer of materials such as sediments, nutrients, and contaminants; and production by plants.

A. Decomposition of Organic Matter and Recycling of Nutrients

Decomposition releases elemental carbon and nutrients (nitrates, phosphates) to the environment with the result that plants and animals that rely on these nutrients can thrive. Additionally, decomposition ensures that wastes and dead organic material do not build up in bodies of water. Terrestrially produced leaf and woody litter and dead aquatic vegetation and plankton sink to the bottom of lakes, streams, and wetlands. Once deposited within fresh waters, this organic matter can be broken down directly by bacteria and fungi. Often however, the material first must be altered by invertebrates, who fragment or shred it before ingesting it. The shredding activities of invertebrates are important because smaller fragments of dead leaves and other organic matter are decomposed by microbes much faster than larger pieces. Additionally, through their feeding activities, invertebrates can enhance the abundance and reproduction of the microbes, which further acts to stimulate the rates of decomposition.

B. The Uptake and Transfer of Materials

Movement of water, sediment, and organic material within and from freshwater systems has a profound effect on global biological, geological, and chemical processes. Freshwater invertebrates may alter the rate at which water flows through sediments. For example, burrowing oligochaetes, nematodes, midge larvae, and crustaceans can increase the rate at which water moves through lake and stream sediments and to nearby groundwaters. Instead of increasing water flow and percolation through sediments, some freshwater invertebrates bind sediments by producing mucus "biofilms" or dense mats of tubes or filaments that reduce water flow and the movement of sediments. This may act to improve water clarity so that sunlight penetrates deeper and plant production is enhanced. Similarly, water clarity may be enhanced by dense populations of bivalves that filter water as they feed and remove suspended matter. Freshwater invertebrates also may influence the concentration of contaminants in the water or bottom sediments by accumulating the contaminant in their bodies or via direct degradation of toxic materials.

C. Production by Plants

In freshwater ecosystems, the synthesis of organic matter by plants depends on the presence of sunlight and nutrients. Rates of primary production are extremely variable across freshwater ecosystems, ranging from low in poorly lit groundwaters or turbid lakes to quite high in well-lit shallow wetlands. The effect that invertebrates have on plant production is indirect but significant. The availability of nutrients for aquatic vegetation and phytoplankton depends to a great extent on the decomposition of organic matter in aquatic sediments and the movement of nutrients. Grazing and movement of invertebrates stir up the bottom sediments and mix water, increasing the availability of nutrients to plants and phytoplankton. Furthermore, planktonic invertebrates that graze on phytoplankton excrete forms of phosphorus and nitrogen that are immediately available to further enhance algal growth. Grazing may be particularly important in the removal of senescent algae, thereby increasing light as well as nutrients.

IV. THREATS TO FRESHWATER FAUNA

The loss of freshwater invertebrate biodiversity because of damage generated by human activities is an immense global problem. A major difficulty in assessing the dimensions of the problem is the poor state of knowledge of freshwater invertebrate taxonomy and distribution and the lack of reliable monitoring to identify loss of diversity. In spite of these problems, it is clear that invertebrate species in some places already have high levels of population decline and extinction.

The major threats to freshwater biota in general, and freshwater invertebrates in particular, are habitat degradation and loss, pollution and reduced water quality, altered hydrologic regimes, and invasion by exotic species. In many situations, the major threats do not act alone but can act synergistically. Thus, for example, streams in agricultural areas may lose their invertebrate fauna due to habitat loss by sedimentation, pollution by pesticides, altered flow regimes due to irrigation, and the introduction of exotic plants and fish. A fifth threat is overexploitation. For example, harvesting of the mussel *Margaritifera margaritifera* for pearls has reduced populations by 90%.

A. Loss and Depletion of Habitat

Humans have damaged water bodies by river channelization and dredging, sand and gravel extraction, wetland drainage, lake and river shore development, dam and barrier construction, water diversion, and levee bank construction. Human activities on land can have indirect but significant effects on fresh waters. For example, sedimentation from poor land use, destruction of riparian vegetation, loss of surface runoff, and loss of water by groundwater extraction for irrigation all can damage fresh waters.

Temporary wetlands are common in many regions. They may have surface water only occasionally, but when they do they are very productive and harbor a distinctive invertebrate fauna adapted to survive long dry periods. Unfortunately, in many areas the importance of temporary wetlands as distinctive habitats of high biodiversity value has not been recognized and many of these wetlands have been plowed, drained, and filled in.

Sedimentation of streams and wetlands from erosion can be caused by activities in watersheds such as land clearance, plowing and tillage of soil, road building, and logging. The mass delivery of sand and silt into water bodies can reduce invertebrate diversity and change species composition by filling in pools in streams, wetlands, ponds, and even lakes by burying porous coarse sediments with layers of fine sediments with very low permeability and by covering aquatic plants. The deep streambed of running-waters has a rich and distinctive fauna dependent on the high permeability of the gravels and sands. Sedimentation can clog the pores in the streambed reducing water movement and oxygen availability, which leads to the loss of fauna. Many streams that originally had stable, habitat-rich channels and a rich invertebrate fauna have been damaged when channels were filled by fine shifting sediment that consequently has led to a greatly depleted fauna.

A property vital to the nature of rivers, and essential for maintainence of invertebrate biodiversity, is connectivity—the unimpeded movement of water, longitudinally from source to mouth, and laterally between the channel and its floodplain. Loss of connectivity greatly alters invertebrate species composition and may reduce diversity.

Dams on rivers are barriers to longitudinal connectivity because they disrupt the downstream movements of nutrients and organic matter and prevent movement of invertebrates and fish. Levees are barriers to lateral connectivity because they are designed to stop lowland rivers from flooding their floodplains; however, periodic floodplain inundation is essential for maintenance of the biodiversity of the river system.

B. Pollution and Reduced Water Quality

Pollutants can enter fresh waters through point sources such as sewage outfalls, or via diffuse nonpoint sources such as runoff from agricultural fields. Pollutants in sewage and agricultural runoff include organic matter and nutrients. Organic pollution of rivers usually is released from point sources, and downstream from the point source, dissolved oxygen concentrations may drop greatly. The normally diverse fauna of the river is eliminated and replaced by an abundance of a few species that can tolerate low oxygen levels. As the organic matter is processed, the river may recover further downstream. In lakes contaminated by organic pollution, decomposition of the organic matter and elevation of the nutrient levels also may create low oxygen conditions in the benthic regions so that the normally diverse benthic invertebrate fauna is replaced by species similar, if not identical, to those found in the anoxic sections of organically polluted rivers.

Nutrients such as phosphorus and nitrogen enter fresh waters either from sewage or from diffuse sources, such as agricultural fertilizers. Nutrients may encourage growth of undesirable algae and may cause algal "blooms" and periods of low oxygen availability, especially in the deeper benthic regions. This process is called cultural eutrophication and is a major problem for lakes in both urbanized and rural areas. In such lakes, the invertebrate fauna, both planktonic and benthic, is depleted in species diversity, but abundances of a few tolerant species may be very high.

Combustion of fossil fuels can generate sulphur and nitrogen oxides that combine with water to form acids in the atmosphere. Acids dissolved in water vapor can move long distances in the atmosphere before entering water bodies as acid rain. This process is a dramatic example of nonpoint source pollution in which pollutants generated in one region exert their effects in another, often distant, area. Many poorly buffered lakes in Europe and North America have become strongly acidic due to acid rain. In acid lakes ($pH < 4.5$), some invertebrate groups such as crustaceans and molluscs are eliminated along with gill-breathing insect larvae. The most acid tolerant taxa appear to be predatory water bugs.

In many parts of the world, irrigation may reduce river flows and create extensive salinization (salt contamination). Salinization occurs when water evaporates from irrigated fields and leaves behinds the salts that were dissolved in the water. The increased salt levels in the soil kill terrestrial vegetation. The concentrated salt can enter rivers and lakes where it eliminates most of the freshwater invertebrates, except for a few tolerant midge larvae and crustaceans. The Aral Sea in Russia was once brackish water and fauna rich. Its freshwater inputs were diverted for irrigation, and it now receives very saline waters from irrigated lands. As a consequence, it has shrunk greatly in size to become a saline sea with a very impoverished fauna.

Many pollutants, such as heavy metals and chlorinated hydrocarbons, are not readily broken down or deactivated. Once they enter a freshwater system through point or nonpoint sources, they persist and continue to cause damage. Heavy metals, such as zinc, copper, cadmium, and lead, may enter a water body from mining operations and poison the system, thereby eliminating many invertebrate taxa. Unfortunately the contamination may persist for many years, long after mining operations ceased. In Wales (Britain), Australia, and many parts of the United States, rivers have remained contaminated 35 to 90 years after mining ceased and they continue to have a depauperate invertebrate fauna.

C. Altered Water Regimes

In the past century with the growth in human populations and the rapidly rising demands for water for human consumption, industry, hydropower, and irrigation, there has been an enormous expansion in the number and scale of dams, barrages, and diversion works. Dams and their large upstream impoundments break the connectivity of rivers. This disruption has been recognized in situations where valuable fish stocks have been threatened. Fishways, porterage schemes, and stocking have been used to reduce threats to fish, but nothing has been done to alleviate the effects on invertebrates. Thus dams may disrupt essential migrations, such as those of shrimps and freshwater crabs in the tropics, and prevent the normal movements of invertebrates down the river.

The waters impounded behind dams may flood valuable habitats—floodplain wetlands, river canyons, even lakes. In Tasmania, Australia, a large dam-created impoundment flooded an entire, isolated lake with a remarkably large beach. This lake, Lake Pedder, contained five endemic invertebrate species, four of which appear to have become extinct. Impoundments behind dams may have a greatly depleted littoral (shoreline) fauna due to rapid changes in water level incurred as a result of dam operations. Drawdown of lake levels to meet hydroelectricity or irrigation can leave many invertebrates stranded. In unimpounded rivers, organic matter and nutrients are transported downstream to be used by downstream invertebrates. This longitudinal transport ceases with impoundment. Water released from dams may come from the deep hypolimnion and consequently may be cold, deoxygenated, and low in organic material—conditions quite unfavorable for riverine invertebrates.

The purposes for which a dam is operated can have dire consequences for downstream invertebrates. Most drastic of all is the simple loss of water. Many dams are water diversion structures that leave the river channel below the dam waterless or with a small minimum flow. In many places, large rivers are left with so little water that they no longer reach the sea. Needless to say the loss of invertebrates is considerable. Dams may be built and operated to generate hydroelectricity, in which case water levels in the channel below the dam fluctuate greatly and quickly as water is released or suddenly cut off to meet electricity generation demands. Alternation between flood and drought creates disturbance levels that permit only the hardiest of invertebrates to exist in the channel. Dams operated to store water for irrigation, hold water in the wet season (winter), and release it for irrigation in the dry season. Consequently, the seasonal patterns of flow are reversed, and wet season floods are greatly diminished. The lack of water in the wet season and high water levels in the dry season disrupt the life cycles of many invertebrates. For many lowland rivers, the effects of greatly diminished floods have been devastating because inundation of the floodplain has been greatly reduced, resulting in a loss of invertebrates that would normally thrive on the floodplain.

D. Invasion of Exotic Species

Exotic species of plants and animals may reach new localities through deliberate introduction by humans or through accidental introduction during transport. The rate of introductions both between and within continents is rising rapidly. The most obvious introductions into freshwater systems have been plants, both aquatic (e.g., water hyacinth) and riparian (e.g., salt cedar), and fish (e.g., carp). Few invertebrates have been introduced deliberately (e.g., opossum shrimp, crayfish), but many, especially crustaceans and mollusks, have been introduced accidentally.

Introduced invertebrates may exert very strong effects on native species. Exotic invertebrates may be vectors for diseases lethal to native species. In Europe, native crayfish stocks have been decimated by the crayfish plague fungus, *Aphanomyces astaci*, which came with the deliberate introduction of North American crayfish. Exotic species may take up habitat space and outcompete local species. The zebra mussel from the Ponto-Caspian region invaded the Great Lakes and Hudson River and now is common throughout eastern North American from Oklahoma and Minnesota east. It monopolizes habitat space and has greatly reduced native mussels. Zebra mussel populations increase rapidly once introduced and through their immense filtration capacity may greatly reduce planktonic organisms increasing the clarity of the water but reducing invertebrate diversity.

Introduced aquatic plants, such as water hyacinth and alligator weed, may blanket the surfaces of tropical lakes and rivers. Such blanketing may reduce water quality by causing deoxygenation or by reducing photosynthesis of native plankton and aquatic plants. This, in turn, diminishes food resources for native invertebrates. Introduced riparian plants, such as tamarisk or salt cedar, may cause the loss of habitat for native invertebrates by dominating the stream banks and often by reducing water levels in the channel. Other introduced plants, such as eucalypts in Spain and Portugal, may produce dead plant material that is not readily consumed by stream invertebrates, thereby causing them to starve.

Many fish are predators of invertebrates. The introduction of exotic fish may increase predation pressure and deplete native invertebrate populations. Introduced carp have reduced invertebrate populations in floodplain wetlands. In New Zealand, introduced trout in invading streams have depleted or eliminated populations of native fish and native invertebrates. Released from the intense grazing pressure of native invertebrates, algae in such invaded streams proliferate and build up to high levels.

See Also the Following Articles

ENDANGERED FRESHWATER INVERTEBRATES • HOTSPOTS • INVERTEBRATES, MARINE, OVERVIEW • INVERTEBRATES, TERRESTRIAL, OVERVIEW • LAKE AND POND ECOSYSTEMS • RIVER ECOSYSTEMS • WETLANDS ECOSYSTEMS

Bibliography

Allan J. D. (1995). *Stream Ecology: Structure and Function of Running Waters.* Chapman & Hall.

Banarescu, P. (1990–1995). *Zoogeography of Fresh Waters,* Volumes 1–3. Aula-Verlag, Wiesbaden, Germany.

Gibert, J. A., Danielopol, D. L. and Stanford, J. A. (1994). (eds) *Groundwater Ecology.* Academic Press, San Diego, CA

Hutchinson, G. E. (1967, 1993). A treatise on limnology, Volumes 2 and 4. Wiley, New York.

Merritt, R. W., and Cummins, K. W. (1996). *An Introduction to the Aquatic Insects of North America.* Kendall/Hunt Publising, Dubuque, Iowa.

Palmer, M. A., Covich, A. P., Finlay, B. J., Gibert, J., Hyde, K. D., Johnson, R. K., Kairesalo, T., Lake, P. S., Lovell, C. R., Naiman, R. J., Ricci, C., Sabater, F., and Strayer, D. (1997). Biodiversity and ecosystem processes in freshwater sediments. *AMBIO* 26. 571–577.

Scheffer, M. (1998). *Ecology of Shallow Water Lakes.* Chapman & Hall, London.

Strayer, D. L., Caraco, N. F., Cole, J. J., Findlay, S., and Pace, M. L. (1999). Transformations of freshwater ecosystems by bivalves: A case study of zebra mussels in the Hudson River. *BioScience* 49, 19–27.

Thorp, J. H., and Covich, A. P. (1991). *Ecology and Classification of North American Freshwater Invertebrates.* Academic Press, New York.

INVERTEBRATES, MARINE, OVERVIEW

P. J. D. Lambshead* and P. H. Schalk†
**The Natural History Museum, London, and †ETI, Universiteit van Amsterdam*

GLOSSARY

benthic That connected with the bed of the sea or a freshwater lake, river, or stream.

deep sea The seabed and immediately overlying water covered by seas at least 200 m deep.

disturbance A biological or physical factor that impacts a population or community by causing death, reduced reproduction, or increased emigration.

infauna Animals living within sediments.

invertebrate All animals that lack a backbone (vertebrae)—that is, most of the animal kingdom.

pelagic That connected with the water column.

planktonic Organisms of many different phyla that float in and are carried by water masses in the pelagic of the sea or freshwater.

production The increase in biomass of an individual, population or community as it grows by converting energy-food into biomass.

AN OVERVIEW OF MARINE invertebrate biodiversity is complicated because there are two distinct marine domains, the open ocean and the seafloor, which are affected by different ecological processes. The physical characteristics of the world ocean are described here both from the viewpoint of a benthic scientist who studies the animals that live on the seabed and a pelagic scientist who studies the organisms that float or swim freely in the water. The pelagic sections include some consideration of single-celled organisms, including examples capable of photosynthesis, because excluding them would be artificial. The processes that produce modern marine diversity patterns are described and the unique features that make marine diversity different from terrestrial diversity are listed. By convention, both pelagic and benthic fauna are not categorized by a functional ecological approach but by the type of equipment used to sample the organisms, which largely depends on their size. The different size classes are listed and examples given of the organisms included in each group. A major limitation in understanding marine invertebrate biodiversity is the inadequate state of taxonomic knowledge of marine organisms, particularly the small and deep-sea taxa. The implications of our igno-

Encyclopedia of Biodiversity, Volume 3

rance for estimating regional or global diversity are discussed. Open ocean pelagic biogeography patterns are described and examples given of species with commonly occurring patterns. Benthic diversity patterns are divided into large and small scale. The large-scale patterns include latitudinal and bathymetric gradients. The change in both diversity and the dominant faunal groups over evolutionary time from the Cambrian to the present is followed. Finally, the unusual hydrothermal and cold seep fauna that have been recently discovered are described.

I. PHYSICAL CHARACTERISTICS OF THE OCEANS

The world ocean covers more than two-thirds of the surface of the earth and as much as 90% of the habitable volume. It is the largest habitat on earth and about one billion people depend on the oceans for their primary protein needs. Yearly, well over 90 million tons of fish and shellfish are harvested.

Only 10% of the oceans' area could be classed as shallow water and this is mostly located around the continental margins. Seas deeper than 2000 m cover half the earth's surface. The seashore zone, the area covered and exposed by tides, corresponds to the ecological littoral region. Shallow seas make up the sublittoral zone, also called the continental shelf, which ends at the shelf break, usually at around 200 m. The continental slope marks the boundary between the continental and ocean floor crusts. It extends from the shelf break down to about 1500 m, so it roughly corresponds to the ecological bathyal region, 200 m to 2000 m.

The continental slope falls away into the continental rise, which extends down to a depth of about 4500 m. Beyond this is the abyssal plain at depths of between 5000 to 6000 m. The continental rise and the abyssal plain roughly correspond to the ecological abyssal region, 2000 to 6000 m. Other features of the seafloor are rocky seamounts and ridges, which may be of a considerable size, and deep trenches down to 11,000 m or more. Trenches are usually a feature of subduction zones where the seafloor buckles and deepens beneath a continental crust. Deep trenches correspond to the ecological hadal region. The continental rise and abyssal plain tend to be relatively flat. Geological features such as terraces and submarine canyons, however, sculpture the continental slope. Pressure increases down the water column at the rate of one atmosphere per 10 m.

Much of the seafloor is covered in sediments; bare rock is rather rare. Near the coast these sediments are of terrigenous origin and may be coarse (e.g. clay, sands, or pebbles). Terrigenous material may be found down to the continental rise and vast areas of the seafloor consist of clay sediments or, in productive areas, biogenic oozes mainly from diatoms, radiolarians, and Foraminifera.

The temperature of the water declines rapidly with depth until it stabilizes at about 4°C, which takes place at approximately 1000 m, depending on latitude. This phenomenon is known as the permanent thermocline. At greater depths, the temperature declines slowly. Water currents are too complex to be discussed here in detail but, broadly speaking, deep water is formed by the sinking of dense, cold saline water in the Antarctic and Arctic. This dense water spreads out across the world ocean, eventually returning to the surface.

In general, marine water is well oxygenated, often at values near saturation. There are areas where rapid seasonal blooms in the upper water column deoxygenate the water leading to oxygen minimum zones. Oxygen concentration falls rapidly with sediment depth, especially in littoral muds and sediments under shallow enclosed seas.

The open ocean or pelagic realm has no obvious boundaries such as the mountain ranges or large water masses (rivers, seas, or oceans) that are so important in driving speciation on land, although to a certain extent the three oceanic basins—Atlantic, Indian, and Pacific—are partially separated by landmasses. The world ocean is essentially a continuum and the most important environmental factors are temperature and salinity, which translate primarily into latitude and depth. However, despite the lack of obvious barriers, speciation has and does occur in the pelagic environment. The geographic species distribution of modern plankton is the product of the geological history of the oceans and continental barriers, limitations of species in adapting to biotic and abiotic environmental factors, and the degree of organism mobility.

The pelagic realm is divided by depth into zones. Light, temperature, and salinity are responsible for the major vertical biodiversity gradients. In the upper 200 m, the epipelagic, sufficient light is available to sustain photosynthetic processes. This is the primary production zone where light and carbon dioxide is converted into carbohydrates, the principle energy (food) source for all open ocean inhabitants. The mesopelagic zone is the realm of the diurnally (daily) migrating organisms and extends from 200 to 2000 m depth. This is the deepest into the water column that light penetrates. Diunal migration is triggered by light and typically involves moving into shallow waters at night and deeper water layers during the day. Possible reasons for such vertical migration may include hiding in the dark depths from predators during daylight and energy conservation achieved by spending part of the time in deeper, colder water layers, thus allowing organisms to reduce their metabolic rate. The bathypelagic (2000 to 6000 m) and abyssopelagic are the aphotic zones, where there is perpetual darkness (Table I).

II. THE FUNDAMENTAL PROCESSES CONTROLLING MARINE BIODIVERSITY

There are three major differences between marine and terrestrial systems that have an enormous impact on biodiversity. The first is that the oceans are three-dimensional; organisms live, feed, and reproduce at all levels in the water column. The second is that light has low penetration through sea water, so photosynthesis only occurs in the upper water column, the photic zone, so most of the oceanic system has no primary production. The third is that most photosynthesis is carried out by tiny single-celled organisms. Macrophotosynthetic organisms (macroalgae or seaweeds) are rare and tend to be concentrated in a thin zone around the edge of the continents in the littoral and sublittoral regions.

Pelagic organisms in the three-dimensional water column have no structure to "hide" in, either for ambush attacks or defense. Essentially, there are only three possible life strategies. An organism can evolve for speed and agility to capture prey and escape predators, or an organism can evolve as a jelly, offering a poor food return for potential predators. A third strategy is rapid growth and reproduction resulting in huge numbers: schools and swarms offer protection to those within. Many pelagic organisms show a patchy distribution for this reason. The organisms within the sediment, the infauna, exist in a two-dimensional world that has a physical structure within which organisms may shelter. The marine system, with regard to biodiversity, may therefore be divided into two separate domains, the pelagic and the benthic. Some infaunal taxa have a pelagic phase and so have a presence in both domains. We may predict that marine pelagic diversity should be lower than terrestrial diversity given the limited choice of life strategy of pelagic organisms coupled with the active or passive mobility of organisms through the water column. The benthic domain is structurally similar to the terrestrial environment and may be predicted to have a similar diversity with the caveat that the absence of large photosynthetic organisms removes an entire biotope.

TABLE I

The Division of the Pelagic Environment into Depth Zones (measured in meters)

Epipelagic	0	200	The photic zone, primary production
Mesopelagic	200	2000	The realm of diurnally migrating organisms
Bathypelagic	2000	6000	The deep-sea realm, continuous darkness
Abyssopelagic	>	6000	The deep-sea troughs, continuous darkness

The lack of primary production in most of the oceanic realm results in a dependence on an organic flux that originates on land or in the photic zone. The organic flux becomes weaker further away from the continents or deeper into the water column. The weakening flux gradient governs benthic biomass, which declines with depth and distance from the continental shorelines. The Russian grab-sampling programs of Belyaev and colleagues from the 1950s onward established the global distribution of benthic biomass (Fig. 1). Productivity has a major influence on biodiversity so the organic flux gradient also will have an impact on benthic biodiversity patterns.

Physical disturbance is also considered to be a key process controlling biodiversity. Physical disturbance was considered to be highest in the littoral region and to decline consistently with water depth down to the abyssal plain. Recent research has caused science to modify this view, as there is now good evidence for hydrodynamic disturbance and seasonal perturbations right down onto the abyssal plain. But in general, the trend is for physical disturbance to decrease with depth. The main physical disturbance processes that show a depth trend are mechanical effects (current energy and wave action), temperature variation (including exposure in the littoral region), and salinity change. Deep trenches are possibly highly disturbed due to slumping of the trench sides.

III. SAMPLING AND ASSESSING MARINE INVERTEBRATE BIODIVERSITY

A. Benthos

Marine invertebrates are divided into three size classes on the pragmatic basis of the equipment used to collect them (Table II). Megafauna are large, visible animals

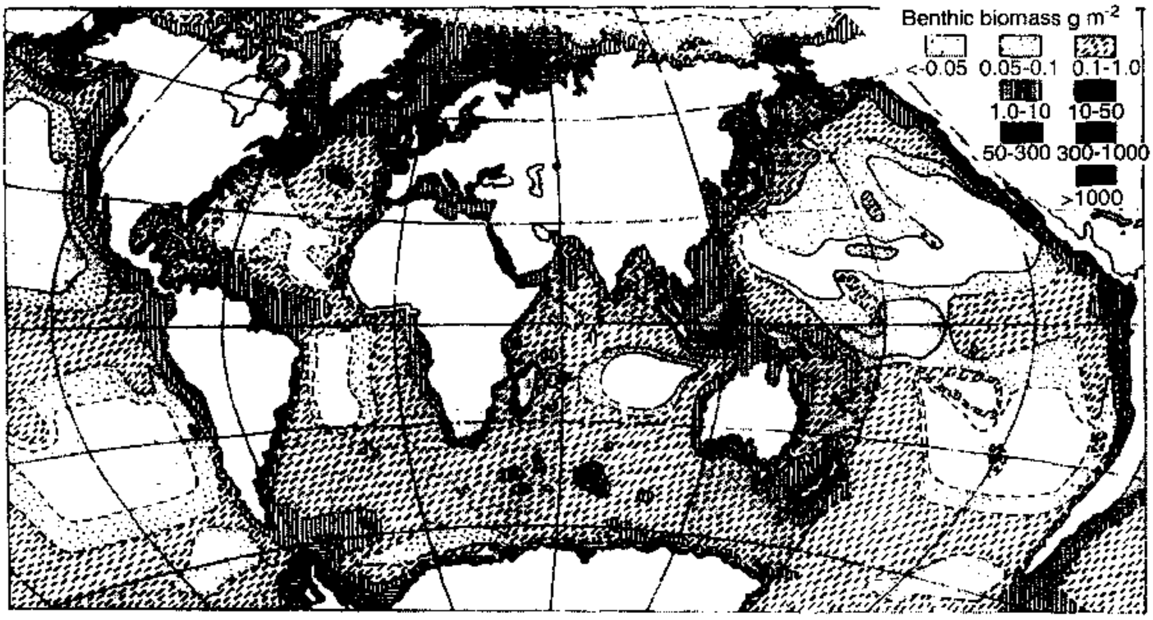


FIGURE 1 The global distribution of benthic biomass in the world ocean based on the work of the Russian sampling programme of Belyaev and colleagues. This work established that biomass declines in samples proportionally to their distance from continents. Biomass is proportional to food input. (From Gage and Tyler, 1991).

that may be seen by eye (or on photographs). Macrofauna are infaunal (sediment dwelling) organisms that are not normally visible but are retained on a 1 mm sieve (0.5 mm or less in the case of deep-sea samples). Meiofauna are the infauna too small to be retained on such a sieve. It is not clear whether these arbitrary classifications also have biological significance. Some taxa are almost entirely in one size class for their entire life history; for example, nematodes are always considered to be meiofauna. Others—for example, polychaetes, although predominantly in one size class, in this case macrofauna—may also have meiofaunal and megafaunal representatives for at least part of their life history. Warwick and colleagues have some evidence, most convincingly from coastal sandy sediments, that the division of organisms into macro and meiofauna is ecologically meaningful. It appears that organisms must be either macrofaunal-sized, to actively burrow through the sediments, or meiofaunal-sized, to slip between the sediment particles, but mechanical limitations prevent them being of intermediate size.

Megafauna are numerically dominated by the phylum Echinodermata, although the Crustacea are also important. The most abundant and diverse macrofauna group are the polychaete worms (Annelida), while meiofaunal nematodes (Nematoda) dominate the benthos, at least in terms of abundance and diversity of meiofauna. Arthropods are also important in both the macro and meiofaunal size classes. The average marine benthic infaunal invertebrate is a soft-bodied worm (Fig. 2), but vermiform animals are less important components of the megafauna.

Littoral benthos from sediments is sampled by inserting a corer into the deposits and washing the removed sediment over an appropriately sized sieve. Sampling offshore is more problematical. Megafauna are collected by equipment such as Agassiz trawls, anchor dredges, or epibenthic sledges towed along the bottom. Such equipment has problems collecting consistent samples and is suitable only for producing qualitative or semiquantitative data. Initially, the smaller infaunal animals were collected with grabs, and these are still used where circumstances dictate but they are inefficient. The development of the box corer by Hessler and Jumars was a critical improvement in the accurate quantitative sampling of macrofauna for biodiversity studies. Development of quantitative meiofauna samplers, notably the Scottish Marine Biological Association's multiple corer, similarly revolutionized investigation into offshore meiofauna biodiversity.

Marine samples taken with modern corers provide data that are more quantitatively accurate than most terrestrial studies, and this has influenced the type of biodiversity questions asked by marine scientists. Ma-

TABLE II

Examples of Marine Benthic Invertebrate Taxa Classified into the Three Size Groups

Size class	Megafauna	Macrofauna	Meiofauna
Collecting equipment	Agassiz trawl, epibenthic sled, anchor dredge	Grab, corer (notably the USNEL box corer)	Corer (notably the SMBA multiple corer)
Examples of taxa	Ascidiacea (sea squirts)	Arthropoda (amphipods, crustaceans, cumaceans, isopods, mites, tanaids)	Nematoda (thread worms)
	Asteroidea (sea stars)		Copepoda (mainly harpacticoids)
	Brachiopoda (lamp shells)	Mollusca (aplacophorans, bivalves, chitons, gastropods, scaphoda)	Gastrotricha Kinorhyncha
	Bryozoa (moss animals)		Loricifera
	Coelenterata (sea firs, jellyfish, sea pens, sea fans, sea anemones, corals)	Oligochaeta ("earth" worms)	Ostracoda
		Phorinda	Tardigrada (water bears)
	Crinoidea (sea lilies, feather stars)	Pogonophora (tube worms)	Turbellaria (flat worms)
	Crustacea (amphipods, barnacles, crabs, prawns shrimps)	Polychaeta (segmented sea worms)	
		Priapulida (worms)	
	Echinoidea (sea urchins)	Nemertea (ribbon worms)	
	Echiura (worms with elongate bifid proboscis)	Sipuncula (peanut worms)	
	Hemichordata (acorn worms)	Turbellaria (flat worms)	
	Holothurioidea (sea cucumbers)		
	Ophiuroidea (brittle stars and basket stars)		
	Pogonophora (including vestimentiferans)		
	Porifera (sponges)		
	Pycnogonids (sea spiders)		

rine research has focused on alpha, or ecological, diversity, which is commonly analyzed using diversity indices that incorporate some measure of evenness as well as species richness. The two most common indices employed are the Shannon-Wiener information function (H′) and, in the deep sea, the Sanders rarefaction method expressed as a curve or index that predicts an expected number of species per number of individuals in a sample. Terrestrial biodiversity analysis tends to focus on larger scale measures of diversity based on species richness per area so it can be difficult to compare published marine diversity patterns with those obtained for terrestrial environments.

B. Pelagos

Functional classification of pelagic organisms is based on locomotion, size, and trophic level or ecosystem function. Plankton are generally passively drifting organisms carried by water movements; examples include bacteria, algae, and small animals. Nekton are actively swimming organisms and so can to a certain extent migrate in the horizontal plane; examples include squid, fish, and some crustaceans. Phytoplankton, marine plants, are primarily unicellular and responsible for primary production. They are the main food source for all life in the open ocean. Zooplankton are animals with a planktonic life style.

Division of pelagic organisms into size classes is related to retention by the different mesh sizes of nets and filters used for sampling (Table III). Pica and nanoplankton are usually collected with remotely operated opening and closing bottle devices, allowing control of the depth zone sampled. To collect organisms sized from microplankton upward, various types of net sampling gear are used ranging from simple ring nets, which

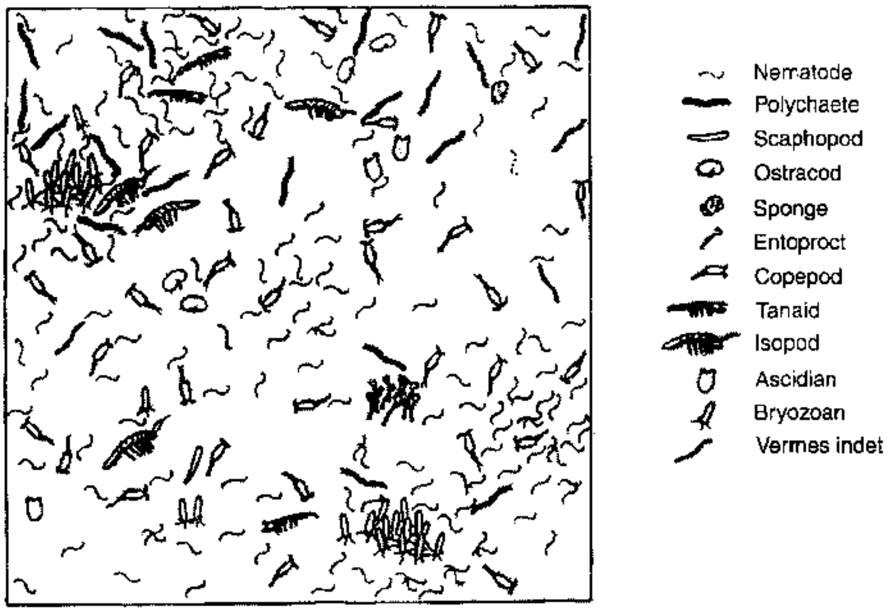


FIGURE 2 The invertebrate animals found in a $0.25\ \text{m}^{2}$ layer of sea sediment from the central North Pacific by Hessler and Jumars. A worm shape is the most common body form, nematodes being particularly abundant. Note that the size of the animals is exaggerated so that they may be seen. (From Gage and Tyler, 1991).

are hauled vertically, to complicated remotely operated opening and closing net systems that can sample horizontal depth layers (e.g., RMT net, Omori net). Nekton are sampled with large, pelagic fish trawls. Often pelagic samplers are used in combination with other equipment that measures physical factors such as depth, temperature, salinity, and light.

Due to the vastness of the pelagic realm relatively little has been studied, although there has been reasonable geographic coverage. It is estimated that less than 5% of its huge volume has been sampled once! Deep-sea and very small planktonic species are particularly undersampled; undoubtedly many new species wait to be discovered.

TABLE III

The Size Classes of Pelagic Organisms with Typical Examples

Planktonic	Picoplankton	<2 μm	Bacteria
	Nanoplankton	2–20 μm	Bacteria, algae, protists
	Microplankton	20–200 μm	Algae, heterotrophs
	Mesoplankton	0.2–2 mm	Copepods
	Macroplankton	> 2 mm	Shrimps, fish larvae, jellyfish
Nektonic	Micronekton	> 2 cm	Small fish, squid, crustaceans
	Nekton	> 5 cm	Fish, squid, marine mammals

One of the main problems with sampling the benthic and pelagic realm is that scientists are literally grabbing in the dark. In contrary to terrestrial research, where the scientist has a visual concept of the environment and can generally "see" what he or she is doing, marine biologists are completely dependent on remotely operated equipment. This has hampered progress in forming concepts of the marine environment as an ecosystem and estimating its biological diversity. Remotely operated vehicles (ROV's) with cameras and sampling devices have improved our understanding of the marine environment significantly in the past decade. The choice of sampling gear has a large impact on the representativeness of the samples for the community as a whole, hence the frequent discoveries of new species or new information about known species. Giant squids present a nice example of a species that have eluded biologists.

IV. TAXONOMY

Estimating, let alone listing, to the number of described species on the earth is a difficult task. For some groups of organisms, usually terrestrial, the actual number of species is reasonably well known with most species named and described; examples include birds, mammals, and higher plants. Other groups have not been comprehensively researched and accurate estimates for species numbers are not available. This category includes beetles, mollusks, and nematode worms. The current best estimate given by the Global Biodiversity Assessment is that there are 1.75 million species known to science. The estimate for the total number of existing species on earth is approximately 12 to 13 million although some scientists think that it may be as high as 20 million.

Classifying and naming the 1.75 million known organisms is an enormous task, but it is vital for comparing and exchanging biodiversity information between different organizations and countries. A profession of approximately 5000 experienced taxonomists worldwide is performing this documentation task. The Kingdom of Plants is the best known and that of Protists least. For the Kingdom Animalia only 20,000 new names are described yearly, including 17,000 species descriptions. The natural history collections stored and maintained in various countries are an important archiving tool for research, and they provide the community at large with a systematic insight into species diversity.

Inventorying marine invertebrates poses exceptional problems. In an area of survey twice as large as the terrestrial environment, marine biology has to deal with problems of the difficult accessibility of the deep sea and the vagueness of the three-dimensional biogeographical borders of the pelagic realm. This makes stock assessments and the definition of distribution ranges extremely problematical.

Species richness is unevenly distributed over the various kingdoms and phyla. The Kingdom Animalia represents the greatest numerical diversity, with well over 1.3 million known species. The terrestrial order of Insecta, in the phylum Arthropoda, takes up about 70% of the known animal diversity.

In both terrestrial and marine domains, the highest diversity is found in the equatorial region. The productivity (in grams carbon per m^2) is more or less equal: 2.2 g C/m^2yr in tropical rain forests and 2.5 g C/m^2yr in coral reefs. In the aquatic environment, current data suggest that coral reefs are the most species-diverse regions and that they are comparable with tropical rain forests. Although reefs only form 0.2% of the seafloor by area, they harbor over 25% of all known marine species. The deep-sea is also believed to be highly diverse. Due to the difficulties in accessing this environment relatively little is known and the discovery of many new species is expected in the coming years (Table IV).

It is often surprising for nontaxonomists to discover that most species in the world have never been described and most probably will never be described. This is especially true for the smaller size classes and marine infauna in general. Marine species have received less attention from systematists than terrestrial organisms,

TABLE IV

Known and Estimated Species Diversity of Various Taxa, including Global Biodiversity Assessment (GBA)

	Described species (×1000)	Estimates (highest) (×1000)	Estimates (lowest) (×1000)	GBA estimate (×1000)	Accuracy of estimate
Viruses	4	1,000	50	400	Bad
Eubacteria & Archae	4	3.000	50	1.000	Bad
Fungi	72	2,700	200	1.500	Reasonable
Protozoa	40	200	60	200	Bad
Algae	40	1,000	150	400	Bad
Plants	270	500	300	320	Good
Nematodes	25	100.000	100	400	Bad
Arthropods					
Crustaceans	40	200	75	150	Average
Arachnids	75	1,000	300	750	Average
Insects	950	100.000	2,000	8.000	Average
Mollusks	70	200	100	200	Average
Chordates	45	55	50	50	Good
Others/diverse	115	800	200	250	Average
Total	1.750	111.655	3.635	13,620	Bad

less than 15% of described species are marine. It is estimated that there are at least 200,000 known animal species and about 50,000 known plant (algae) species in the oceans. Fish (26,000 species) and marine mammals (250 species) belong to the best known taxa.

Most marine animals live in the deep sea or are tiny and were out of sight and out of mind. Only in the second half of the 20th century has the equipment become available to comprehensively collect deep-sea and small planktonic organisms, and it is still an expensive business limited to a handful of professionals. When this is compared with the long and vigorous tradition of both professional and amateur taxonomic study in the terrestrial domain, the backwardness of marine invertebrate systematics is understandable.

The taxonomically best known marine area in the world is the coastal strip along northwest Europe and to a lesser extent the coasts of the rest of Europe and North America. Outside these zones, less than half the marine invertebrate species are known to science. For example, a recent study of 16 box cores taken from bathyal depths off the British coastline revealed 304 polychaete species, of which only 17% could be associated with a known species. A study of 6 liters of coral reef sediment from Hawaii found 158 polychaete species, of which only 30% had been previously described. The largest study of macrofauna ever undertaken in the deep sea in the bathyal, northwest Atlantic counted 798 infaunal species from 233 box cores, but less than half could be named (Table V). This study illustrates the importance of polychaetes. In general, at least half the macrofauna species of any deep-sea sample will be new to science.

TABLE V

The Species and Family Diversity of the Major Invertebrate Macrofauna Phyla in Box Cores from the Bathyal Northwest Atlantic

Taxa	Number of families	Number of species
Annelids	54	398
Arthropoda	40	185
Mollusca	43	106
Echinodermata	13	39
Nemertina	1	22
Cnidaria	10	19
Sipuncula	3	15
Hemichordata	1	4
Echiura	2	4
Priapulida	1	2
Brachiopoda	1	2
Ectoprocta	1	1

TABLE VI

Species Diversity per m^2 of a Mudflat from a British Estuary for Some of the Major Meiofauna Groups

Taxa	Number of species
Polychaetes	5
Nematodes	40
Oligochaetes	2
Copepods	7
Ostracods	2
Hydroids	1
Kinorhynchs	2

The systematics of meiofauna is even less well known. For example, a nematode study in the Arctic Ocean reported that only 4% of the 92 species found were known to science, while a study in the Venezuela Basin could only name 1.5% of the 136 species found. The taxonomic breakdown of different meiofauna groups in a single region has not been properly explored due to the difficulty of gathering the disparate taxonomic expertise. Some taxa, such as turbellarians, may be important but are commonly ignored in samples because they are so difficult to identify (Table VI).

The identification of marine benthic invertebrates is problematical not just because of lack of taxonomic coverage but because the taxonomic state of the art is inadequate. For many of the soft bodied groups, the species concept (i.e., the criteria used to define a species) is poorly defined. For example, the apparently well-known opportunistic polychaete species *Capitella capitata* and *Streblospio benedecti* have been found to be a complex of genetic sibling species (15 in the case of *C. capitata*).

The lack of species descriptions of small marine animals makes it difficult to assess regional diversity, and this is another factor that encourages marine biologists to consider diversity only over small scales. Only marine megafauna have an adequately censused taxonomy.

V. GLOBAL DIVERSITY

It is impossible accurately to assess the total number of species in the oceans because of limited sampling, the nonrandomness of sampling locations, and inadequate taxonomy. The best estimate has come from Grassle and Maciolek, who sampled benthic macrofauna for 176 km along a bathymetric gradient 1.2 km to 2.0 km deep in the bathyal zone of the northeast Atlantic. They discovered a rate of species turnover of one new species detected every km. Assuming one new species per km^2, given that there is 10^8 km^2 of deep sea, this implies that there may be up to 10^8 (100,000,000) species in the seas. Such extrapolation techniques are notoriously inaccurate so this figure should be treated with caution but Grassle and Maciolek's calculation is surprisingly similar to the estimate obtained for insect diversity of rain forests by Erwin who used substantially the same technique, suggesting that marine benthic and terrestrial diversity may be similar. A much better knowledge of the geographic ranges of the smaller benthic organisms is essential for an accurate estimation of global marine diversity. This will only be possible when some of the taxonomic problems surrounding these organisms are resolved.

The marine environment certainly demonstrates a greater diversity at higher taxonomic levels than the terrestrial domain. Life evolved in the oceans and many phyla never evolved to survive in the harsher terrestrial environment. It is highly likely that new metazoan phyla remain to be discovered, especially among the small organisms in marine sediments. The free-living phylum Loricifera was only described in 1983 and the commensal phylum, Cycliophora, was first found in 1995 (Table VII).

The distribution of benthic invertebrate species richness around the global ocean basins has not been comprehensively assessed. Schopf has shown that both ectoprocts and bivalve mollusks are twice as diverse in the Pacific than the Atlantic. This may reflect the relative sizes of the two oceans, or it may be the product of the relative age of the basins; the Atlantic is younger than the Pacific.

There is still some controversy over whether the highest marine benthic invertebrate diversity is found in coastal regions or the deep sea. Certainly the deep sea is a much larger environment than shallow water, so it might be expected to be home to more species; but it is unclear whether it is more diverse per area sampled. Originally, the low abundance of animals found in the deep sea caused scientists to assume that the highest diversities must occur on the coastal shelf. However, in the 1960s Saunders found the deep sea unexpectedly diverse, indeed with a higher species richness than shallow water. A series of studies appeared to confirm this view but recently an analysis has reopened the debate. Gray carried out a similar study to that of Grassle and Maciolek (noted earlier) but on the Norwegian coastal shelf. He found a similar, or even higher, rate of accumulation of species with sampling on the shelf compared to the deep sea.

TABLE VII

Comparison of Free-Living Invertebrate Diversity of Terrestrial and Marine Environments at a High Taxonomic Level

Endemic marine phyla	Endemic terrestrial phyla	Cosmopolitan phyla
Brachiopoda	Onychophora	Annelida
Chaetognatha		Arthropoda
Ctenophora		Chordata
Echinodermata		Cnidaria
Echiura		Ectoprocta
Gnathostomulida		Gastrotricha
Hemichordata		Kamptozoa
Kinoryhncha		Mollusca
Loricifera		Nematoda
Phoronida		Nemertea
Placozoa		Platyhelminthes
Priapula		Porifera
		Rotifera
		Sipuncula
		Tardigrada

VI. PELAGIC DIVERSITY PATTERNS

Van der Spoel and Heijman summarized the general distribution pattern types and biogeographic regions in the open ocean for phyto- and zooplankton species as Arctic, Subarctic, Cool-Temperate, Warm-Temperate, Tropical, Temperate and Subtropical, Subantarctic, and Antarctic. These regions principally reflect seawater temperature and latitude. Contrary to earlier beliefs, relatively few plankton and nekton species have a true cosmopolitan circumglobal distribution. Despite the lack of obvious barriers in the oceans, there is a surprising variety of the types of species distribution. Species specific characteristics and ecosystem relations play an important role in combination with abiotic factors in determining distribution patterns, but many are further modified by regional influences such as ocean basins, currents, and divergence and convergence zones. Typical and easily recognizable biogeographic patterns are belt-shaped patterns related to latitudes (temperature regimes) and neritic distributions (basically around the continental coasts and shallow water areas). The distri-

TABLE VIII

Pelagic Distribution Patterns

Type	Region	Example species
Circum-global	Cosmopolitan	*Stylocheiron maximum*
Belt-shaped patterns	Cosmopolitan	*Rhizosolenia alata*
	(Sub) tropical	*Clausocalanus paupulus*
	Central Waters	*Stylocheiron suhmi*
	Temperate N and S	*Sagita plantonis f zetesios*
	High latitude N	*Calanus glacialis*
	High latitude S	*Thysanoessa macrura*
Central water patterns	Atlantic, Indian, Pacific	*Euphausia brevis*
	Indo-Pacific	*Hydromyles globulosa*
Endemic patterns	Atlantic	*Euphausia americana*
	Indian	*Desmopteris gardineri*
	Pacific	*Sagitta pseudoserratodentata*
Neritic patterns	Cold water north	*Gammarellis homaris*
	Cold water south	*Euphausia crystallorophias*
	Warm water	*Charybdis smithli*
Bathypelagic patterns	Deep water	*Cyclothone pseudopallida*

bution around the Antarctic continent of *Euphausia superba* (the shrimp commonly known as krill) is an example of a neritic distribution (Table VIII).

Many widely distributed pelagic plankton species show North-South or East-West variation. Without active migration, genetic contacts between populations from one end of the range to the other are limited or even nonexistent, promoting genetic drift. In addition, the differing environmental circumstances (abiotic and biotic) at the ends of each range can give rise to phenotypic differences. This is reflected by both morphological and ecological variation within a species, a first step toward the development of new species. An example of North-South variation is found in the shell shape and size of many species of Pteropoda (pelagic Mollusca). East-West variation can be found for example in *Eucalanus subtenuis* (copepod, Crustacea).

It is expected that over coming decade, knowledge of pelagic biogeography and open ocean distributions will develop exponentially (Table IX).

VII. BENTHIC LARGE-SCALE DIVERSITY PATTERNS

A. Latitudinal Gradients

Latitudinal gradients of terrestrial taxa commonly display a decline in species richness from the equator to

TABLE IX

Species Diversity in the Atlantic (A), Pacific (P), and Indian (I) Oceans for three example groups: Chaetognatha, Pteropoda, and Euphausiidae (largest diversity is found in the equatorial region)

	Chaetognatha			Pteropoda			Euphausidae		
	A	P	I	A	P	I	A	P	I
Arctic	1	1	—	3	4	—	5	5	—
Subarctic	9	6	—	20	17	—	10	10	—
40°N–40°S	25	34	29	120	90	95	25	30	30
Subantarctic	9	8	8	17	17	17	10	10	10
Antarctic	4	4	4	11	11	11	5	5	5
Total	48	53	41	171	139	123	55	60	45

the poles. Many different processes have been suggested to explain such gradients including competition, predation, mutualism, parasitism, and host diversity, but the most convincing explanation is that the diversity gradient follows the gradient of solar energy. There are two potential mechanisms that might cause this relationship. One suggestion is that latitudinal gradients in solar radiation cause concomitant gradients in productivity and hence diversity. The other suggestion is that the relationship between solar energy and diversity is the result of increased evolutionary speed in warmer conditions. The explanations are not mutually exclusive.

The relationship between latitude and diversity for marine benthic coastal fauna could best be described as confusing, with different studies and different taxa yielding conflicting results. For example, if the diversity of core samples is compared using diversity indices, then no significant difference is found between tropical and temperate samples. This experiment has been repeated for both macrofauna by Warwick and colleagues and meiofauna by Boucher and Lambshead.

However, Roy and colleagues were able to test for a latitudinal diversity gradient for prosobranch, gastropod mollusks using data based on number of species per degree of latitude. Unusually for marine studies, this data is similar in scale and format to the terrestrial diversity data used for analyzing large-scale patterns and it is noteworthy that it produces convincing latitudinal diversity gradients (Fig. 3). Furthermore, the prosobranch diversity gradient follows the solar energy gradient in a similar way to terrestrial latitudinal gradients.

The deep sea is an intriguing location to search for latitudinal gradients because solar energy can have no direct impact on deep-sea diversity. The deep sea is uniformly cold and relatively stable so it can be argued that temperature cannot be a factor in controlling diversity though influencing evolutionary rates. Furthermore, productivity gradients in the deep sea are the result of gradients in the food flux to the seafloor, which vary according to a number of processes.

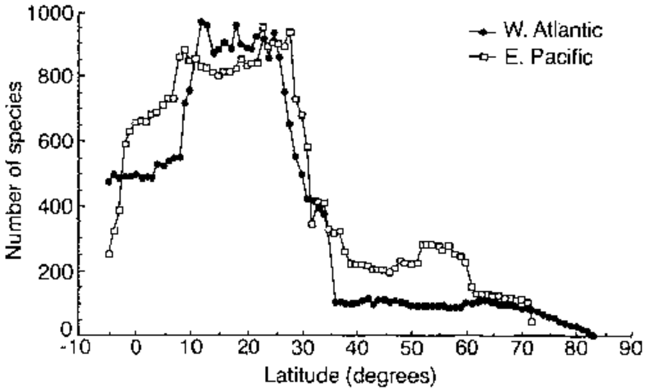


FIGURE 3 The number of American coastal species of prosobranch gastropods per degree of latitude. On both coastlines the diversity peaks at 10 to 30°N. (From Roy *et al.*, 1998).

Rex and colleagues found that deep-sea gastropod mollusks, bivalve mollusks, and isopods showed a decline in diversity from the equator to the Norwegian Sea in the North Atlantic (Fig. 4). The explanation is not obvious, as productivity tends to increase northward.

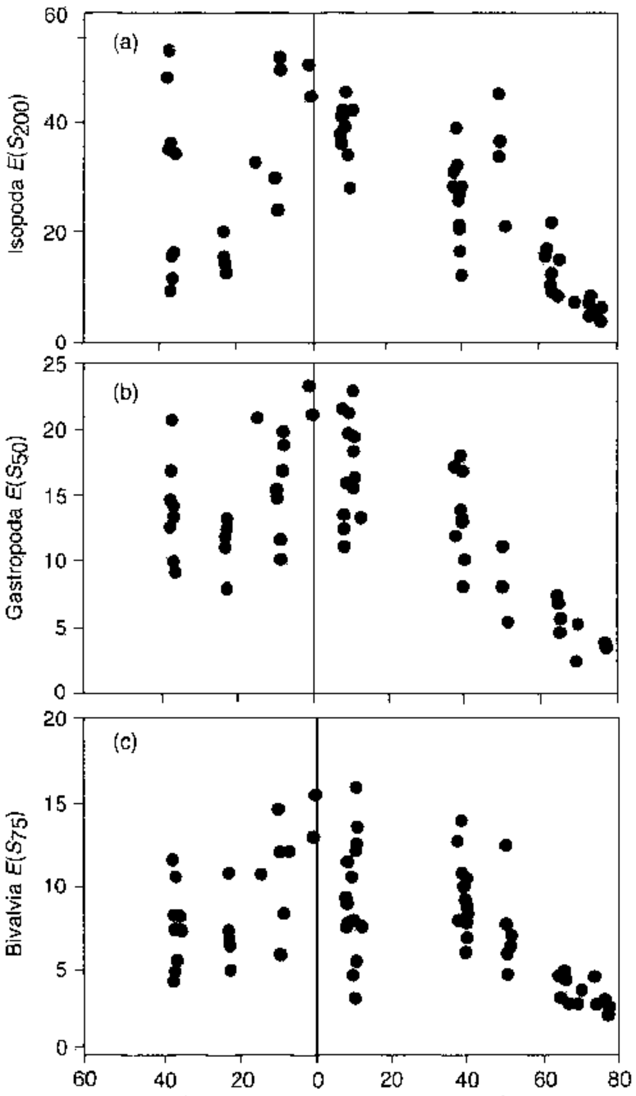


FIGURE 4 The change in species diversity with latitude for Atlantic deep-sea (a) isopods, (b) gastropods; and (c) bivalves. Rex and colleagues demonstrated that species richness for these taxa falls from the equator northward into the Norwegian Sea. Species richness has been calculated using the Sanders rarefaction index. (Reprinted by permission from *Nature* 365, 636–639 © (1993) Macmillan Magazines Ltd.)

Similar gradients have been searched for in the South Atlantic by Brey and colleagues but have not been found, so it is possible that the North Atlantic gradient is due to some individual feature of this ocean, possibly the Quaternary glaciation.

Analysis of the smaller invertebrates by Lambshead and colleagues has shown different patterns. Nematode diversity appears to be correlated with the food flux to the seafloor (i.e., a productivity gradient). In the North Atlantic, there is a weak increase in diversity northward. In the Pacific, nematode diversity declines northward away from the equator, as does the food flux to the seafloor (Fig. 5).

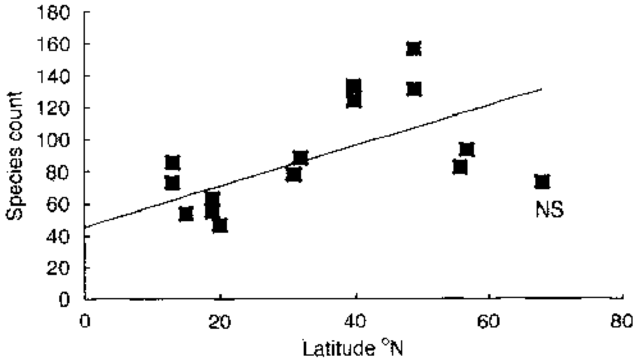


FIGURE 5 The count of the number of nematode worm species per station in the North Atlantic plotted against latitude. There is a gradual increase in species richness following the increase in food supply from the equator northward. This pattern extends as far north as the Norwegian Sea (marked NS) where there is a marked drop in species richness. The pattern of diversity being positively linked to food input appears widespread for small invertebrates.

B. Bathymetric Gradients

Bathymetric gradients are probably better understood than any other large-scale diversity pattern for marine benthic invertebrates and for once a consistent pattern is found in all size classes. Rex showed that a number of macrofaunal taxa followed a parabolic diversity curve with depth, diversity peaking at about 2 to 3000 m (Fig. 6). A number of studies have since confirmed this finding. A similar pattern was found for megabenthos by Haedrich and colleagues and for nematodes by Boucher and Lambshead (Fig. 7).

The exact peak in each bathymetric diversity gradient varies with different taxa and with different loca-

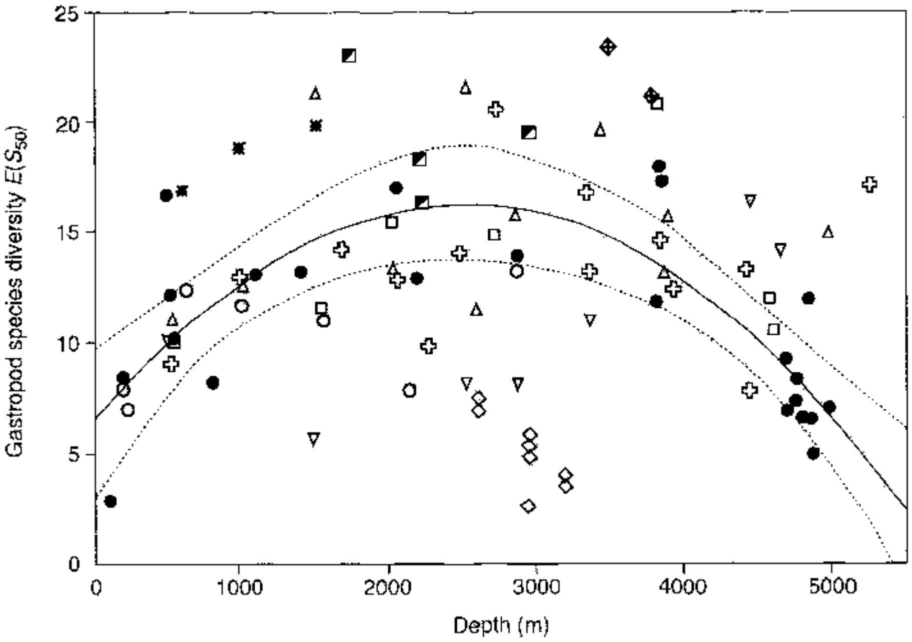


FIGURE 6 The relationship between gastropod diversity, as measured by the Sanders rarefaction index, and depth in 85 epibenthic sledge samples from various locations in the Atlantic Ocean and Norwegian Sea. Rex and colleagues showed that diversity peaks at bathal depths declining in both shallow and deeper sediments. The pattern is probably caused by the effects of disturbance and food availability on diversity. Note that the regression lines and 95% confidence limits are for the samples from the North American Basin only. (From Ormond *et al.*, 1997).

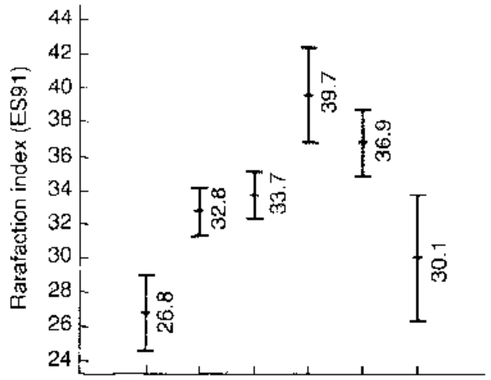


FIGURE 7 The species richness of marine nematodes in sediments from six biotopes, temperate estuaries, tropical shallow water, temperate shallow water, bathyal depths, abyssal depths, and hadal depths. Species richness is measured using the Sanders rarefaction index. Diversity is highest in bathyal sediments.

TABLE X

The Species Diversity of Nematode Samples from Different Depths and Habitats from the British Isles Westward into the Deep Sea

Location of sample	Depth m	Species richness	Number of individuals studied
Clyde Estuary	High water, beach	20	272
Clyde Estuary	Mid water, beach	30	1268
Clyde Estuary	Low water, beach	56	549
Irish Sea	56	46	150
Rockall Trough	545	45	104
Rockall Trough	835	50	102
Rockall Trough	1474	56	127
Porcupine Abyssal Plain	4840	71	493

tions and, for all we know, there may be temporal variations. However, the basic pattern of diversity peaking in the bathyal or upper abyssal region appears consistent. This consistency suggests that similar processes are responsible for all taxa and locations. Two obvious explanations are that both productivity and disturbance decline with depth. Computer models (e.g. those produced by Huston) have demonstrated how the interaction of productivity and disturbance gradients can theoretically give parabolic diversity patterns.

The exact species composition of the benthic community changes rapidly with depth for some taxa with species having a preferred depth range. Other taxa have wide depth ranges. This has been better studied for the larger organisms than for small organisms due to taxonomic reasons. A wide variety of reasons probably explain the different strategies displayed by different taxa but, in general, the sharpest discontinuity is found at the shelf break with a second boundary at about 1000 and 2000 m. So the basic pattern is of zones of gradual change in species composition separated by areas of rapid change.

VIII. BENTHIC SMALL-SCALE DIVERSITY PATTERNS

One of the most intriguing observations about marine benthic invertebrates is the high species richness found in individual core samples (Table X).

The observation that a large number of similar species can be found together in samples of apparently featureless sediment has always represented a challenge for "equilibrium" explanations for the ecological processes that control diversity. Such explanations depend on the hypothesis that coexisting species must avoid competition by each possessing some unique adaptational specialization. This hypothesis is difficult to accept for such a simple organism as a free-living nematode worm, many of which seem to have similar functional requirements.

An alternative explanation is that large numbers of species can coexist because the sediment is divided into transient patches by biological and physical small-scale disturbance and uneven distribution of the vertical food flux. Grassle and colleagues have coined the phrase *spatio-temporal mosaic* to describe this process (Fig. 8). A new patch is initiated by a disturbance event, or a food fall, and colonization to exploit the new situation increases diversity. As the patch ages and is exploited, competition increases and diversity drops. But at some point, a new event will recreate a new patch. In this theory, the sediment consists of a heterogeneous mosaic of unique patches of different ages and histories. For any given species, there is a suitable habitat within dispersion distance in the mosaic, and when that patch becomes unsuitable, another opens up nearby. The theory depends on limited large-scale disturbance and a lack of barriers for dispersion so it is considered most applicable to the deep sea (and possibly rain forests!).

This spatio-temporal mosaic theory predicts that areas of the deep-sea subject to a flux of phytodetritus would have a more patchy distribution of animals and higher species diversity than other locations. This has

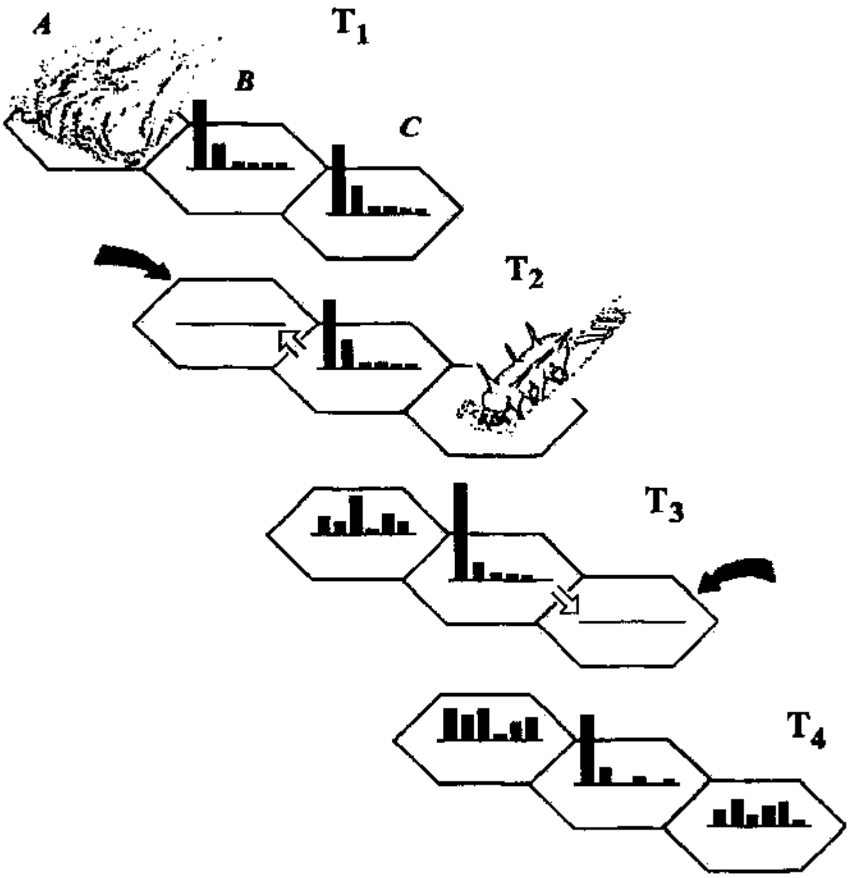


FIGURE 8 The spatio-temporal mosaic theory of Grassle and colleagues. The concept is that the sediments consist of a series of patches, each of which evolves through an unpredictable cycle. The example shows three patches, A, B, and C through times T_1. T_2, T_3, and T_4. Each vertical bar is a species and the hight of the bar represents the abundance of that species. At T_1, patch A is physically disturbed removing all the animals by T_2. The patch is recolonized from the water column (dark arrow) and by migration from adjacent patches (light arrow). By T_3 and T_4, the patch is diverse with six species, five of which are equally abundant. Patch B is undisturbed through time. As the patch matures and is exhausted, the diversity drops to four species though competition. One of these species becomes completely dominant. Initially, patch C is similar to patch A, but it is biologically disturbed by the feeding behavior of a large organism at T_2. This empties the patch allowing recolonization, forming a diverse community. This theory explains how apparently featureless marine sediments can support such a high diversity of life. For any species there is always a suitable new patch available nearby when it exhausts or is outcompeted in its current location. (From Paterson).

been tested for nematode worms by comparing the small-scale spatial distribution and species diversity of nematodes from a station in the phytodetritus-enriched Porcupine Abyssal Plain with a station from the more oligotrophic Madeira Abyssal Plain. Nematodes from the enriched site were more diverse and more species showed aggregation. Interestingly, the species were not aggregated in concordance together around the food or evenly randomly aggregated with respect to each other but showed discordance (i.e., species tended to aggregate in different places). This is consistent with Grassle's theory.

IX. DIVERSITY OVER EVOLUTIONARY TIME

The marine fossil record is more complete than the terrestrial because the ocean floor tends to accumulate sediment. Even so, only certain fauna, notably those with "hard parts," will form an adequate fossil record. The dominant macrofauna group, the polychaetes, presents a limited fossil record, and the dominant meiofauna group, the nematodes, has no record at all. The available fossil record suggests that marine species have an existence of about 4 million years. This indicates on average a 25% species turnover per million years. Assuming that the oceans contain 10^7 (10,000,000) species, we might predict that 2.5 species would become extinct due to natural processes every year. In practice, the fossil record suggests that there are periods of relative quiescence followed by mass extinction and subsequent speciation. The fossil record also implies that 95% of all the marine species that have ever existed are extinct.

The history of marine biodiversity is best tracked at the family rather than the species level to remove some of the noise associated with an incomplete fossil record. This does tend to make the mass extinction events appear less dramatic. For example, only 54% of marine families were lost in the Permian extinction event, but it is estimated that 77 to 96% of all species became extinct.

The early Cambrian shows a rapid increase in diversity that tends to flatten out in the middle and late Cambrian (Fig. 9). Most animal phyla appear in the record in this phase. The Cambrian diversity includes a number of "archaic" forms such as trilobites, hyoliths, and inarticulate brachiopods that decline after the Cambrian period. Diversity is not high in the Cambrian and rather unspecialized detritus and low suspension feeding organisms functionally dominate communities, suggesting a simple ecology.

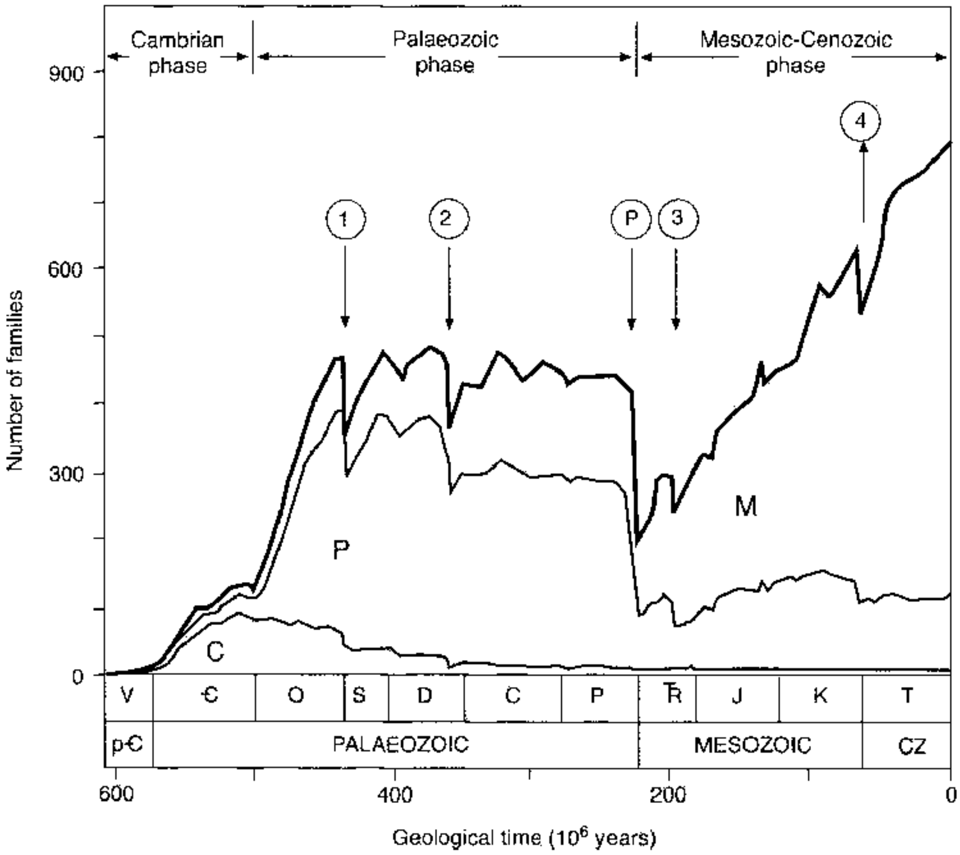


FIGURE 9 The family diversity of marine animals as measured by their fossil record from 600 M years ago to the present day. Graph C shows the rise and fall of the diversity of the Cambrian fauna of unspecialized detritus and low suspension feeding organisms, P show the rise and fall of the diversity of the Palaeozoic fauna of sessile benthic organisms, and M shows the rise of the diversity of the modern fauna of sediment movers and shell breaking predators. Note that sampling error may have exaggerated modern diversity. The arrows indicate periods of mass extinction, 1 is at the end of the Ordovician, 2 in the late Devonian, 3 at the end of the Triassic, and 4 at the end of the Cretaceous (the dinosaur extinction). P indicates the massive Permian extinction. (Graph from Systematics, Ecology, and the Biodiversity Crisis, ed. by N. Eldredge, © 1992 Columbia University Press. Reprinted with permission of the publisher.)

The Ordovician sees a steep rise in diversity but the curve then flattens for some 200 million years. Periods of mass extinction are detectable, notably at the end of the Ordovician and in the late Devonian, but in general diversity stabilized. This new Palaeozoic diversity was associated with an evolutionary radiation of sessile benthic organisms such as crinoids, articulate brachiopods, stenolomate bryozoans, and tabulate and rugose corals.

The Palaeozoic period ended with the catastrophic Permian-Triassic mass extinction. Diversity then apparently rose steadily, with a small extinction at the end of the Cretaceous, past the levels achieved in the Paleozoic until the unique peak of the present day. Current family diversity is apparently twice as high as the Palaeozoic stable level. The Palaeozoic community never seems to have recovered and the modern high diversity is associated with a new fauna of shell-breaking predators and sediment movers. This fauna includes the familiar sea urchins, bivalve and gastropod mollusks, and crustaceans such as shrimps and crabs.

A variety of explanations have been given for this pattern. A number of authors have pointed to the association between the rise in taxonomic diversity and a rise in functional diversity. This almost certainly explains the Ordovician increase in diversity, which is essentially because of increased exploitation of marine sediments. Explanations for the Mesozoic-Cenozoic explosion in diversity are more controversial. One argument is that the breakup of the Pangaea supercontinent caused increased diversity through increased climatic variation leading to more endemicity. It is not clear how this argument would apply to the deep sea, where the bulk of marine diversity is to be found. Another suggestion is that evolution in the exploitation of the terrestrial domain has increased nutrient runoff into the oceans concomitantly increasing marine diversity. A similar argument has combined these two explanations suggesting that volcanism associated with continental breakup increased nutrient flux into the oceans.

Finally, Raup has suggested that the high modern marine diversity is an artifact of the fossil record. He observed that across the Mesozoic-Cenozoic time periods there are more younger rock available for study than older rock. If Fig. 9 is adjusted to allow for this, then there is a slower recovery from the Permian mass extinction and modern diversity peaks at much the same level as in the Palaeozoic. If true, it implies that there is a maximum global carrying capacity for marine biodiversity and that the Permian-Triassic extinction was even more devastating than Figure 9 suggests since the new graph would show that the Palaeozoic benthic fauna has declined consistently since the extinction event. This is possible as the new Mesozoic fauna included many organisms that would tend to disrupt the stable sediment environment on which the sessile Palaeozoic fauna depended.

X. HYDROTHERMAL VENTS AND COLD SEEPS

The special biological communities around hydrothermal vents were first discovered in 1977 on the Galapogos Rift at a depth of 2500 m. They are now known to be widely distributed in the oceans at tectonically active sites such as subduction and fracture zones, ocean-floor spreading centers, and back-arc basins. Vents are formed where sea water penetrates through fissures in the ocean floor deep into the earth's crust. The water becomes heated and escapes back to the surface through hydrothermal vents. The water temperature at a vent varies from mildly warm, 23°C, to superheated, 350°C and can be rich in sulphide and metalliferous ions. The hottest vents are called "smokers" from the precipitation of minerals in the water.

Cold seeps that release sulphide and methane-enriched water have been recorded from a number of areas in the Atlantic and Pacific Oceans. Similar chemical conditions have been found around whale carcasses that are oil rich.

These sulphide-rich habitats are small and ephemeral. The evidence is that many have a life span that measures in decades, although cold seeps may last longer. However, they do appear to be commonplace and may be thought of as chains of islands across the seafloor, separated by distances of 1 to 100 km.

Vent communities have a noticeably different biodiversity from the surrounding ocean floor. The fascinating feature of vent communities is that they derive their energy from chemosynthetic primary production, by reducing compounds such as hydrogen sulphide, rather than photosynthetic sources. It is not clear whether these communities are entirely chemically independent of the photosynthetic world but certainly they are energetically independent.

Biomass is high at hydrothermal vents and primary productivity may be double or triple that of the overlying water. Dense mats of bacteria are found, notably *Thiomicrospira*. A number of invertebrate species feed in dense colonies either directly on these bacteria or by means of symbiotic chemoautotrophic bacteria. Evolutionary selection pressure has been for large, fast-growing species at these productive but ephemeral sites.

Conversely, species diversity is much lower at vents than at other benthic habitats, illustrating once again Rosenzweig's "paradox of enrichment." The dominant species vary from place to place and from depth to depth. For example, the shallow vents off the Palos Verdes Peninsula in California are dominated by the black abalone, *Halioctis cracherodii*; the Mid-Atlantic Ridge vents are characterized by two species of caridean shrimp; and the Eastern Pacific Rise and Galapagos Spreading Centre vents are dominated by the red-plumed, tube dwelling vestimentiferan worms (notably *Riftia pachyptila*) and the large bivalves *Calyptogena magnifica* and *Bathymodiolus thermophilus*. The fauna around the vents are often found to be new to science. Other fauna include a number of polychaetes, crustacea (notably including a number of decapods, and a new primitive genus of barnacle, which is a relic of the Mesozoic).

More than 150 new species, 50 new genera, and 20 new families or subfamilies have been identified with sulphide-rich vents and seeps. Endemism is considered to be high and it is noteworthy that many of the fauna do not appear to be particularly well adapted for high dispersal. Genetic studies on *Bathymodiolus* have shown strong genetic similarity between organisms at sites only 8 km apart but high dissimilarity at sites separated by 2200 km. Vents that are distinctly separated from the chain of sulphide-rich habitats such as the Mariana Trough or Hawaiin volcanic seamounts do show considerable faunistic differences from the norm and the Atlantic sites are less diverse than the Pacific.

See Also the Following Articles

ENDANGERED MARINE INVERTEBRATES • INVERTEBRATES, FRESHWATER, OVERVIEW • INVERTEBRATES, TERRESTRIAL, OVERVIEW • MARINE ECOSYSTEMS • PELAGIC ECOSYSTEMS • PLANKTON, STATUS AND ROLE OF • VENTS

Bibliography

Butman, C. A., Carlton, J. T., and many other authors. (1995). *Understanding Marine Diversity*. National Academy Press, Washington, DC.

Cullen, V. (Ed.) (1995). *Marine Biodiversity I. Oceanus*. Woods Hole Oceanographic Institution, Woods Hole.

Gage, J. D., and Tyler, P. A. (1991). *Deep-Sea Biology*. Cambridge University Press, Cambridge.

Heywood, V. H. (Ed.) (1995). *Global Biodiversity Assessment*. UNEP. Cambridge University Press, Cambridge.

Lambshead, P. J. D. (1993). Recent developments in marine benthic biodiversity research. *Océanis* 19, 5

Ormond, R. F. G., Gage, J. D., and Angel, M. V. (Eds.) (1997). *Marine Biodiversity, Patterns and Processes*. Cambridge University Press, Cambridge.

Roy, K., Jablonski, D., Velentine, J. W., and Rosenberg, G. (1998). Marine latitudinal diversity gradients: tests of causal hypotheses. *Proc. Natl. Acad. Sci. USA*, 3699–3702.

Spoel, van der, S., and Heijman, R. P. (1983). A comparative atlas of zooplankton. Biological patterns in the oceans. Bunge, Utrecht. 1–186.

INVERTEBRATES, TERRESTRIAL, OVERVIEW

Olof Andrén
Swedish University of Agricultural Sciences

GLOSSARY

Annelida Segmented worms, a relatively species-poor phylum.

Arthropoda Joint-legged animals, the most species-rich phylum existing today.

enchytraeids Segmented worms that are related to but smaller than earthworms.

invertebrates Animals without a backbone (vertebrae), ranging from unicellular Protozoa to multicellular, complex organisms such as insects.

phylum Highest level of taxonomic division in the animal kingdom, followed in descending order by class, order, family, genus, and species.

THE TERRESTRIAL INVERTEBRATES CONTAIN THE MAJORITY OF THE WORLD'S ANIMAL SPECIES, although marine environments harbor more phyla representing more widely separate animal groups. Only comparatively few of the phyla have managed to adapt to terrestrial conditions, that is drought, ultraviolet radiation, no support of water against gravity, as well as rapid temperature fluctuations. However, some of the phyla that have adapted to the terrestrial environment have been very successful from a biodiversity standpoint, particularly the insects (phylum Arthropoda, class Insecta). When the basic demands for terrestrial life are fulfilled (water management, etc.), the terrestrial environment is well suited for the evolution of new species. It contains many areally defined habitats, for example, deserts, mountains, and forests, as well as vertical strata, for example, the air volume, treetops, hilltops, valleys, bushes, grasses, the litter layer, topsoil, and subsoil—each with a different evolutionary pressure.

I. THE TERRESTRIAL INVERTEBRATES

A. Taxonomic Groups, Fully Adapted to Terrestrial Conditions

The truly successful and well-adapted terrestrial invertebrates belong to only two phyla, Arthropoda (joint-legged animals) and Mollusca (snails, etc.). Other taxa, such as Protozoa, Aschelminthes (nematodes, etc.), and Annelida (earthworms, etc.), are limited by their need of an aquatic or at least water-saturated environment, such as in the litter layer and in soil, and a lack of protection against ultraviolet rays. These groups can only move in the true terrestrial environment during nights with favorable weather conditions.

The Insecta are the largest class of animals. About one million species are described, but there are possibly as many as five million species present in the world today. The bulk of the insect species are in the orders Coleoptera (beetles), Lepidoptera (butterflies and moths), Hymenoptera (wasps, ants, bees), and Diptera (flies) (Table I and Figs. 1–3). Other Arthropoda that successfully cope with the terrestrial environment include the classes Diplopoda (millipedes) and Chilopoda (centipedes) and the similarly built but smaller and less known Symphyla and Pauropoda.

In the class Arachnoidea, scorpions (order Scorpionidea) and spiders (order Araneae) may be the most obvious. However, the smaller mites and ticks (order Acari) are a highly diverse group, containing parasites, predators, fungivores, and detritivores, and are common in almost all litter layers and soils (Fig. 4).

TABLE I

Insect Orders, Common Names of Typical Members, and Approximate Numbers of Described Species

Order	Typical members	Number of species
Protura	Telsontails	100
Thysanura	Bristletails, silverfish	700
Collembola	Springtails	2000
Ephemeroptera	Mayflies	1500
Odonata	Dragonflies, damselflies	5000
Orthoptera	Grasshoppers, crickets, cockroaches, mantids	23,000
Isoptera	Termites	2000
Plecoptera	Stoneflies	1500
Dermaptera	Earwigs	1100
Embioptera	Webspinners	150
Psocoptera	Psocids	1100
Zoraptera	Zorapterans	20
Mallophaga	Chewing lice	3000
Anoplura	Sucking lice	250
Thysanoptera	Thrips	3200
Hemiptera	Bugs	23,000
Homoptera	Cicadas, aphids, hoppers	32,000
Neuroptera	Dobsonflies, fishflies, snakeflies, antlions	4700
Coleoptera	Beetles	277,000
Strepsiptera	Twisted-wing parasites	300
Mecoptera	Scorpionflies	350
Trichoptera	Caddisflies	4500
Lepidoptera	Butterflies and moths	112,000
Diptera	True flies	85,000
Siphonaptera	Fleas	1100
Hymenoptera	Sawflies, wasps, ants, bees	103,000

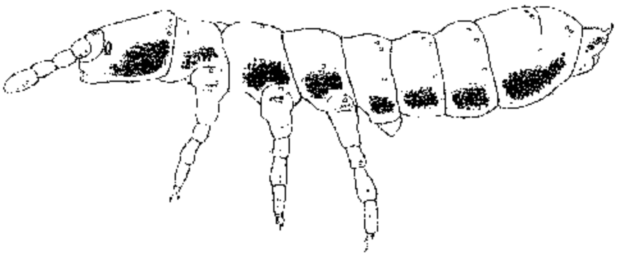


FIGURE 1 The collembolan (springtail) *Protaphorura* sp. (class Insecta, order Collembola). (From Coleman and Crossley, 1996.)

The crustaceans (phylum Arthropoda, class Crustacea) are mainly aquatic, but the terrestrial isopods (order Isopoda, the "pill-bugs") are quite widespread, even in dry climate zones. However, they lack the water regulation of, for example, insects and spiders and are restricted to moist micro-environments and nocturnal habits.

Among the molluscs (phylum Mollusca) only a few classes have terrestrial capability, and the snails (class Gastropoda) have successfully coped with the dry atmosphere by developing a shell into which they can escape.

FIGURE 2 Example of biodiversity among Lepidoptera. (From *Bilder ur naturen. En gåva till goda barn.* P.A. Norstedt & Söner, 1842. Scanned from Dal, 1996, with permission from the author.)

FIGURE 3 An early micrograph (woodcarving) of a flea (class Insecta, order Siphonaptera). The lettering is not explained in the original. (From *Swenska Mercurius*, 1682. Scanned from Dal, 1996, with permission from the author.)

B. Taxonomic Groups with Limited Adaptations

The subkingdom Protozoa contains microscopic animals such as flagellates, ciliates, and amoebae. These are found almost everywhere but are restricted to water films or water bodies. However, several groups have resting stages that can withstand severe desiccation and have the capacity to rapidly reproduce when conditions improve. Thus Protozoa can regulate bacterial biomass and numbers in the soil.

The phylum Platyhelminthes contains primitive flatworms that lack an anus and sometimes even guts. They are mainly parasitic (e.g., tapeworms), but there are free-living species in the class Turbellaria.

Nematodes (phylum Aschelminthes, class Nematoda) are unsegmented, mainly microscopic worms that often have the capacity to form resting stages that can withstand drought and low temperatures. Thus the nematodes have successfully invaded most environments, including arable land (several are pests, but the majority are not) and the inside of other animals (several are internal parasites of mammals, insects, etc.) (Fig. 5).

Segmented worms are in the phylum Annelida, class Oligochaeta, and contain the well-known "ecological engineers"—the earthworms. Well over 1200 species of earthworms have been described. Lesser known but very abundant in boreal forests and wetlands are the smaller enchytraeids, which are also segmented (Fig. 6).

II. THE TERRESTRIAL ENVIRONMENT

The terrestrial environment is harsher than marine or freshwater environments. Shortage of water, ultraviolet radiation, rapidly fluctuating temperatures, and a num-

FIGURE 4 Ventral and dorsal scanning electron micrographs of the oribatid mite *Carabodes* sp. (class Arachnoidea, order Acari), found in boreal forest litter. (Photo courtesy of T. Persson.)

FIGURE 5 Micrograph of a free-living rhabditid nematode (class Nematoda, order Rhabditida). (Photo courtesy of J. Lagerlöf.)

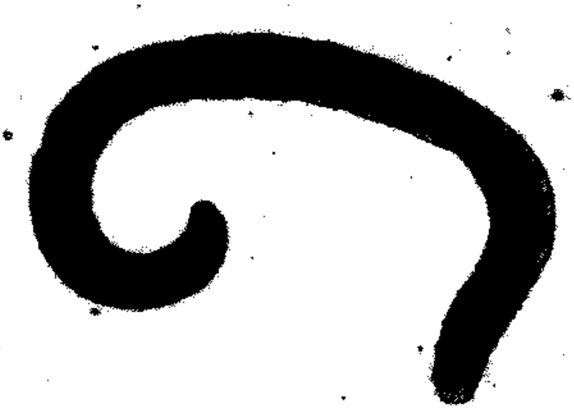


FIGURE 6 Micrograph of a soil-living enchytraeid worm (phylum Annelida, class Oligochaeta). (Photo courtesy of J. Lagerlöf.)

ber of obstacles against movement and/or the dissemination of offspring create survival problems, but also opportunities for evolution and speciation. The fact that most of the known animal species are terrestrial is due to the diversity of habitats and a fairly high probability for isolation of populations—a necessary condition for speciation.

Invertebrate biodiversity ranges from the very low in the polar deserts (a few insect and nematode species) to the extremely high diversity found in some rain forest areas, particularly of insects. Note that the high biodiversity in rain forest may be due partly to the high humidity, which reduces the environmental limitations of animal groups that have limited terrestrial adaptations (e.g., slugs). The various ecosystems and habitats that harbor terrestrial invertebrate diversity are described elsewhere in the Encyclopedia.

See Also the Following Articles

INVERTEBRATES, FRESHWATER, OVERVIEW • INVERTEBRATES, MARINE, OVERVIEW

Bibliography

Andrén, O., Lindberg, T., Paustian, K., and Rosswall, T. (eds.). (1990). Ecology of arable land—Organisms, carbon and nitrogen cycling. *Ecol. Bull.* 40 (Munksgaard International, Copenhagen.)

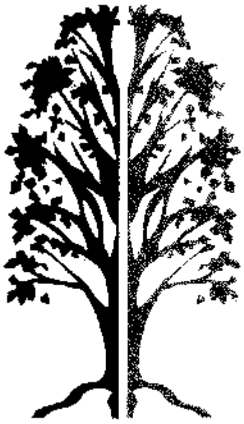
Coleman, D. C., and Crossley, D. A., Jr. (1996). *Fundamentals of Soil Ecology*. Academic Press, San Diego.

Dal, B. (1996). *Sveriges Zoologiska Litteratur* 1483–1920. Orbis Pictus, Kjuge, Sweden.

Dindal, D. (1990). *Soil Biology Guide*. John Wiley & Sons, New York.

Edwards, C. A., and Lofty, J. R. (1977). *Biology of Earthworms*. Chapman & Hall, London.

Stachowitsch, M. (1992). *The Invertebrates. An Illustrated Glossary*. John Wiley & Sons, New York.



ISLAND BIOGEOGRAPHY

Dieter Mueller-Dombois
University of Hawaii

THIS ARTICLE EXPLAINS first the concepts of biogeography as a science concerned with organismic patterns of distribution and abundance on our planet. Island biogeography specifically deals with such patterns in relation to isolated land fragments. The chapter then introduces a botanical and ecological viewpoint by the subject matter of geobotany in tabulated form. Three biogeographic theories are briefly described, which are generally inspirational, especially for the geobotanical approach. This approach contributes particularly to biogeography by a set of proven methods used in vegetation science for plant biodiversity surveys. These methods are outlined as floristic checklist, vegetation sampling, and data processing and display techniques. The latter are illustrated by an island example. This is followed by a new experimental design for coordinated Pacific-wide biodiversity research, which refers to a combination of horizontal and vertical transect approaches. They involve a within biome comparison across the Pacific upland forests, together with vertical transects to study watershed functions down to the coastal and near-shore marine habitats. This program is now being initiated as the PABITRA (Pacific-Asia Biodiversity Transect) net, the Pacific Island Branch of DIWPA (Diversitas in Western Pacific and Asia).

I. CLARIFICATION OF CONCEPTS

A. Island Biogeography

Biogeography is a scientific approach to understanding the distribution and abundance of living things, the biota, on our planet. Island biogeographers are primarily interested in isolated areas and the study of fragmented life zones and their relation to the biota. But what living things or biota are included? Nearly all groups are studied: plants, birds, insects, other animals, humans, fungi, fishes, disease organisms, and so on. From this list it is clear that biogeography is not a single discipline. Instead, it is a unifying principle for scientists of different disciplines. The unifying principle is their interest in the distribution and abundance of the organisms with which they have a greater familiarity. Thus, botanists, ornithologists, entomologists, mamologists, mycologists, and anthropologists can all come together and be unified by their interest of biogeography.

Encyclopedia of Biodiversity, Volume 3

B. Biogeographic Scales

The study of distribution and abundance of biota can be applied at any level of scale in space and time.

1. Spatial Scales

Traditionally, biogeographers were, and still are, interested in biotic patterns occurring at global and intercontinental scales. Geographically speaking, these are small scales that provide for broad overviews on maps or satellite images. Scientists were concerned with distinguishing and mapping large-area distribution patterns, based on landscape physiognomy, such as tropical rain forests, deserts, savannas, and temperate grasslands, commonly called "biomes." A parallel concern was to distinguish broad patterns of species distribution, which could be mapped as "biotic provinces," areas distinguished as different centers of biodiversity with their "own" floras and faunas. Lately, biogeographers have become concerned with distribution patterns occurring at more detailed spatial scales, such as within-archipelago migration patterns, or the distribution of biota along individual mountain slopes. Moreover, the aspect of abundance versus rareness of species and other taxa has become of concern to biogeographers, an area treated traditionally by ecologists.

2. Timescales

Present-day distribution and abundance patterns of biota are usually the result of past events and historical processes. Historical processes that led to present-day distribution patterns of biota are often measured at geological timescales in millions of years. Similarly long timescales are considered for biological evolution. But organic evolution by mutations and hybridization can occur at any timescale. Other timescales of biogeographic interest relate to the concepts of succession following disturbances and to phenological change. Successional change in terms of primary succession, which refers to vegetation development on new geological surfaces, can be considered at long timescales involving hundreds and thousands of years. Primary succession can also be related to soil development (= pedogenesis). Secondary succession, defined as following disturbances on already developed soils or previously vegetated substrates, may be considered at plant-demographic timescales, involving a few years, decades, or hundreds of years, if based on the life cycles of certain long-lived tree species. Phenology relates to seasonal changes in biota. The latter two biotic change-with-time concepts—succession and phenology—were developed from ecological research in biogeography. Thus, biogeography as a research approach to discover chronological changes overlaps with those aspects of ecology that deal with distribution and abundance of biota at successional and phenological timescales and also with evolutionary research that seeks to unveil migration patterns and phylogenetic relationships over long and short timescales.

3. Geobotany

The term *geobotany* is derived from geographical botany, the biogeographical study of plants. In an effort to clarify Central European and Anglo-American terminology in this broad study area, Mueller-Dombois and Ellenberg (1974) synthesized the specializations within the area of geobotany in tabular form, which is reproduced in Table I.

Biogeography was originally understood as consisting primarily of phytogeography and zoogeography. Today, following a review of recent textbooks (e.g., Cox and Moore, 1993; Hengeveld, 1990; Huggett, 1998; and others), one can consider all the ecological disciplines listed under Anglo-American equivalents (Table I) as aspects of plant biogeography.

II. THREE INSPIRING THEORIES

A. The Theory of Island Biogeography

This theory, originally proposed by MacArthur and Wilson (1963, 1967), is practically synonymous with the concept of "island biogeography." It proposes that species equilibria are formed on islands in relation to the size of land area and its distance from biotic source areas. An equilibrium is suggested where the rate of invasion equals the rate of extinction of island biota. These are the intersection points on Fig. 1. Curves sloping from the left to the abscissa represent decreasing rates of invasion of biota, from near to far source areas; curves sloping from the zero point on the abscissa up to the right ordinate represent increasing rates of extinction, from large to small islands. The curves are based on a complicated mathematical model. However, the model is easily understood from Figure 1. For example, a large archipelago, such as the Fijian Islands, situated nearer to continental source areas for the dispersion of biota, would have richer species equilibria than a large archipelago, such as the Hawaiian Islands, which is much farther removed from any biotic source area.

The reality of species equilibria is highly questionable, yet the theory has been and continues to be very inspirational. If applied not simply to species per se, but instead to different life forms, such as indigenous trees, ferns, or shrubs, the results may show differing

TABLE I

Areas of Specialization within the Field of Geobotany, Their Synonyms, and Anglo-American Equivalents[a]

Area of specialization (and synonyms, European terms)	Subject matter	Anglo-American equivalents (and synonyms)
Floristic geobotany	Study of geographic distribution of plant taxa and their evolutionary relationships	Plant geography (phytogeography)
Sociological geobotany (vegetation science, plant sociology, phytosociology, phytocoenology)	Study of composition, development, geographic distribution, and environmental relationships of plant communities	Synecology (community ecology, plant ecology in part)
Ecological geobotany (plant ecology)		
Autecology (ecophysiology)	Study of physiological functions of individual organisms in the field environments and communities; life-history studies of species or ecotypes	Autecology (physiological ecology, population ecology in part)
Demecology (population ecology)	Study of structure and function of populations	Population ecology
	Study of genetic variation in populations	Genecology
Synecology (habitat science; ecosystem research)	Study of habitat factors and the physiological response of species and species groups to these factors; study of community functioning, and niche functions of plant populations in an ecosystem context	Ecosystem ecology (community process ecology, functional ecology, systems ecology)
Historical geobotany	Study of historical origins and development of populations and communities	Paleobotany (paleoecology)

[a] From Mueller-Dombois and Ellenberg (1974).

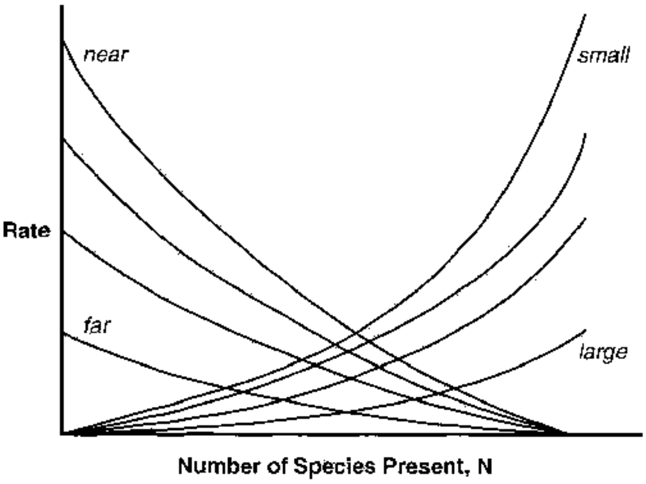


FIGURE 1 The MacArthur and Wilson (1963, 1967) models of dynamic species equilibria as controlled by distance of islands from biotic source areas (from near to far) and by size of islands (from small to large).

values of species richness and endemism in relation to size of island areas and degrees of isolation. Moreover, these in turn may lead, with additional ecological studies, to a better understanding of the function of biodiversity in different island ecosystems.

At this point, another limitation of this theory should be mentioned. This relates to the fact that size of island area is only a most general predictor of species richness. At least elevational range and substrate type should be added to make the theory more predictable. This brings us to the next theory.

B. The Biome Theory

This theory predicts that there are broad life zones that are indicated by groups of biota of key plant life forms, which are controlled by certain broad-area climatic and edaphic (soil) parameters. For example, the biome theory predicts that in mountainous environments there are altitudinal life zones (Holdridge, 1967) or vegetation zones (Mueller-Dombois and Ellenberg, 1974) that can be distinguished by tree species and other life forms into lowland, upland, and high-altitude zones. Also, such familiar terms as *desert, grassland, deciduous forest, coniferous forest,* and *tundra* depict different latitudinal biome types, which in turn can be defined by climatic parameters as "zonobiomes" (Walter *et al.*, 1975). It is assumed that conditions for life within a biome are more homogenous than life conditions across different biomes (such as grassland versus desert).

Applied to island biogeography, the biome theory lends itself to a more appropriate refinement in the analysis of biodiversity than is offered by the two ecosystem parameters—size and isolation—in the above described theory of island biogeography. For example, comparative biodiversity research within a Pacific-wide biome type, such as the montane rain forest on volcanic high islands of basaltic origin, is scientifically more satisfying than biodiversity research based simply on size of island area. The size approach groups different types of islands into the same category, which thus can be a very heterogeneous category that includes atolls, raised limestone, and volcanic high islands. Using the biome theory, environmental gradients can be studied among physically similar islands. Examples of environmental gradients for island research of the same biome type are given in Fig. 2. Atoll and reef islands can be considered as belonging to the same biome, in this case, the same "pedobiome," meaning they have similar, marine-derived, substrates and similar low elevations.

C. The Theory of Succession

Both the island biogeography and biome theories thus outlined contain elements of succession or community and ecosystem development. MacArthur and Wilson (1967) speak of five fundamental processes as the most difficult to study in biogeography. These are listed as (a) dispersal, (b) invasion, (c) competition, (d) adaptation, and (e) extinction of species. In his concept of succession, the early, influential American ecologist Clements (1916) recognized six processes:(a) nudation, (b) migration, (c) ecesis (establishment by reproduction), (d) competition, (e) reaction (habitat change through organisms), and (f) final stabilization, the climax community. All of these processes are of concern in biogeographic research. For example, the process of "dispersal" and "migration" of species among islands and their "adaptation" in terms of speciation were chosen as the main topics in a recent treatment of Hawaiian biogeography (Wagner and Funk, 1995). Other processes, such as the establishment and regeneration of populations (ecesis) after major disturbances (nudations), their "invasion" relative to "extinction" or their "final" assemblages in communities or ecosystems, also fall into the realm of biogeographical research. Clements proposed a final stabilization, called climax, while MacArthur and Wilson proposed a dynamic (final) species equilibrium as explained earlier. Both, the climax and species equilibrium concepts have been severely criticized. Yet dynamic equilibria remain an area of ecological and biogeographical interest because an understanding of dynamic processes is essential for an improved theory of island biogeography.

In connection with our long-term research on the native Metrosideros forest dieback in Hawaii, I introduced a model of ecosystem development based on the theory of succession (Mueller-Dombois, 1986). This is diagrammatically portrayed in Fig. 3. The model addresses the concepts of climax as well as those of primary and secondary succession and habitat change with time in a single island biome, the montane tropical rain forest of Hawaii. Here, the process of nudation can be a new pāhoehoe lava flow or a volcanic ash blanket. Both, the volcanic ash and pāhoehoe substrates, may achieve a "climax" in vegetation development in about 1000 to 5000 years in the rain forest climate. After that, a regression phase sets in very slowly, characterized by cation leaching, increasing occlusion of phosphorus (Crews *et al.*, 1995), formation of secondary aluminum (gibbsite) and iron (goethite) minerals, and advanced desilication (Fox *et al.*, 1991). With time, here estimated as 1 million years, the forest undergoes a number of generation turnovers in the form of canopy breakdown or gap formation and recovery. These may be synchronized over larger or smaller areas, depending on the disturbance regime, the cohort structure, and the aging pattern in the forest mosaic. The different kinds of dieback or canopy failure, depicted on the diagram as demographic events, reflect the habitat changes in terms of soil water and nutrient relations that occur over the long timescale. The model also implies that after the plant biomass or biophilic nutrient climax, forest recovery yields successively less tall Metrosideros forests. Associated with the progressive and regressive phases of ecosystem development, different species assemblages also occur. Species equilibria have not been detected among native species. The question of invasive nonnative species is a separate major problem of island biogeography that is discussed in another chapter of this book.

III. VEGETATION ANALYSIS METHODS

A. Floristic Checklist Methods

"Floras" are lists of plant species. When more elaborate, they are books with species descriptions, keys, and illus-

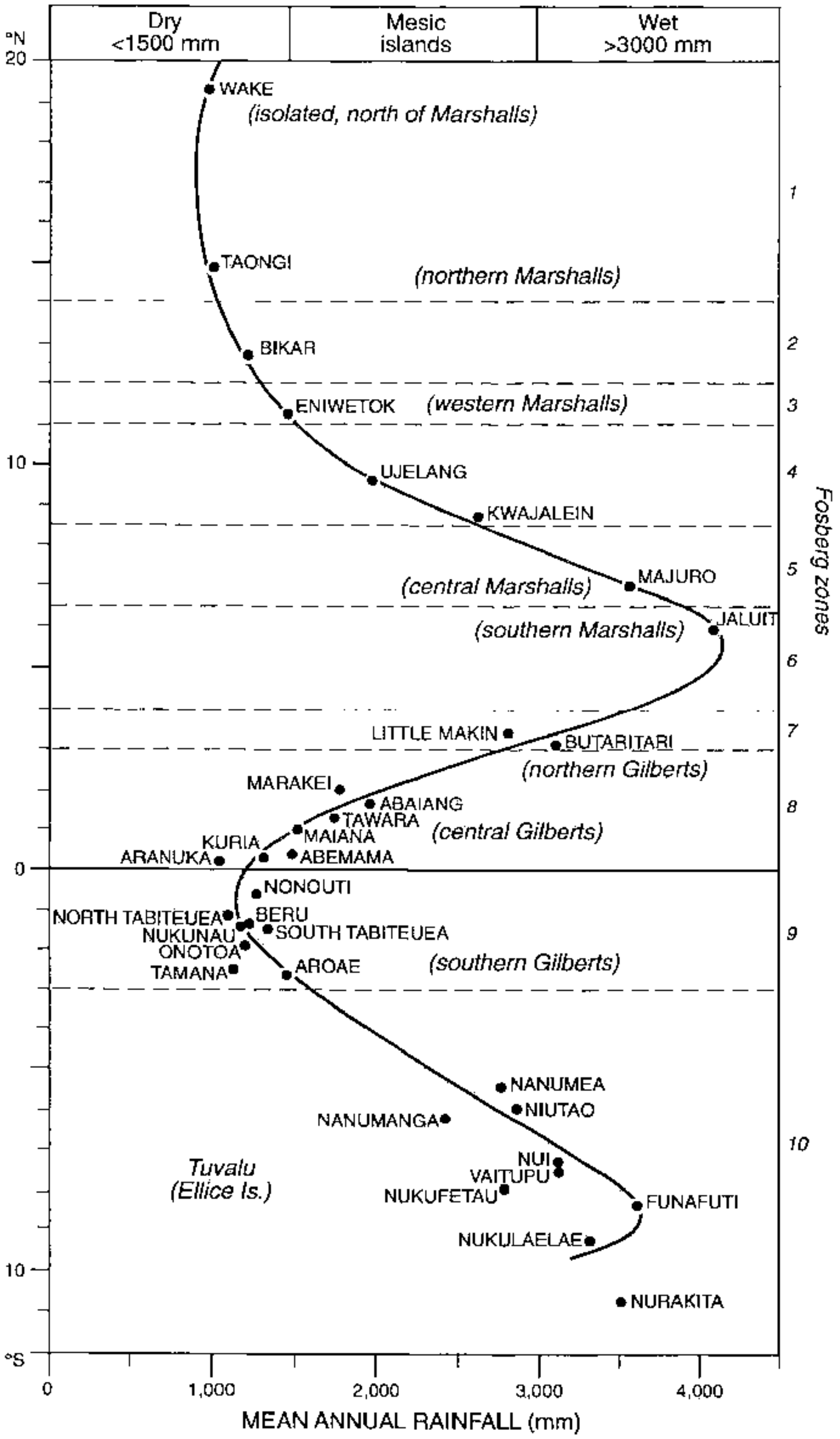


FIGURE 2 Latitudinal gradients of mean annual rainfall in eastern Micronesia, from the northern Marshalls to the southern Gilberts (Kiribati) and Ellice Islands (Tuvalu), summarized by Fosberg zones. Adapted with modifications from Stoddart (1992).

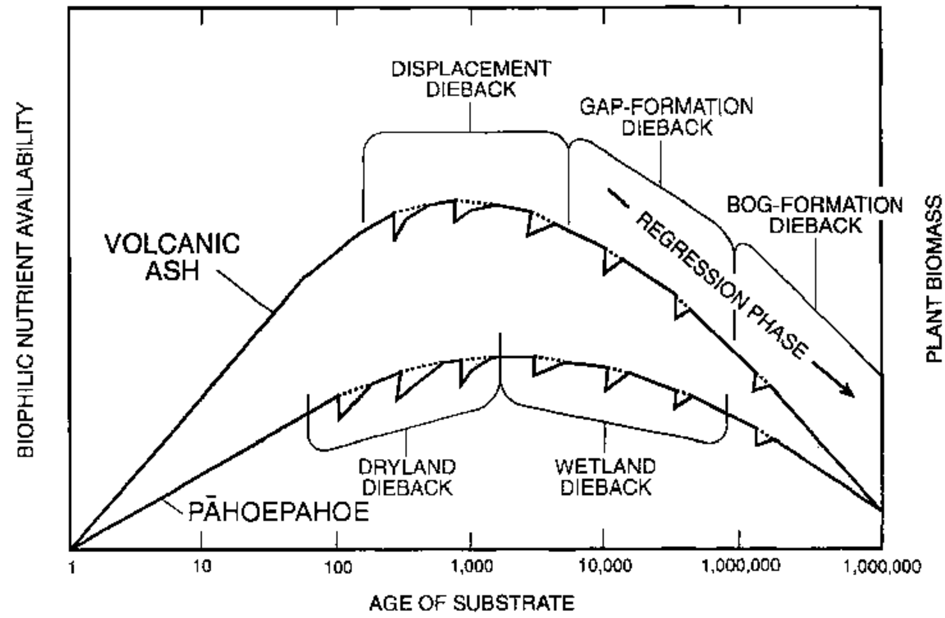


FIGURE 3 A model of primary succession and regression in the Hawaiian rain forest biome with superimposed secondary successions as caused by Metrosideros canopy failures. These are related to demographic growth cycles and the climatic disturbance regime. A "climax" is shown by the maximum in plant biomass and biophilic nutrient availability. After this climax, indicated mainly by the height of Metrosideros canopy trees, the regression phase is caused by leaching of cations, occlusion of phosphorus, and complete desilication. From Mueller-Dombois, 1986.

trations for plant identification. Floras usually give little information on vegetation, which is the most visible component in terrestrial ecosystems, there performing the function of primary producer. "Vegetation" is the plant cover of an area in which species play different roles in terms of their abundance or rareness and in terms of their life forms, life histories, and physiology. In nature, plants often interact with one another in communities. "Plant communities" can be distinguished by differences in life-form structure—such as forests versus shrub- or grassland—or by differences in species assemblages. In the study of island biogeography, floristic checklist methods can be applied at the level of biogeographic provinces, which may serve for floristic comparisons of archipelagoes. They can also be applied among individual islands of the same archipelago. At both these broad levels of scale, tests of similarity versus dissimilarity can be applied from floristic records or subregional floras where these exist. However, field research is needed if one wants to compare the floras within the same biome on different archipelagoes, because this information is usually not available. A simple "walk-through method" can provide this information without formalized sampling designs, such as plots or transects. Such broad floristic surveys also help in familiarizing with the territory and, when used together with aerial photographs and topographic maps, they can result in vegetation maps outlining tentative communities. However, floristic comparisons among communities of the same biome in an archipelago, such as the rain forest across the high Hawaiian islands, require a more formalized checklist method in representative sample plots, such as described by the relevé method (Mueller-Dombois and Ellenberg, 1974, discussed later).

B. Vegetation Sampling Methods

A number of field methods are available. Their proper use depends on the purpose of sampling and the type of vegetation. Here, I will consider plant biodiversity as the general objective of sampling.

1. The Relevé Method

This is a widely applied and proven floristic checklist method, ideal for many, if not most, plant biodiversity surveys. It is based on plots, called relevés (meaning abstracts), whose sizes are based on the "minimal area concept."

a. Determining Relevé Size

A minimal area for vegetation sampling can be established by the "nested plot technique." This technique consists of lining out a small quadrat, 0.5 × 0.5 m, and then enumerating all species encountered in this small quadrat. Next, the quadrat size is doubled into a rectangle of 0.5 × 1 m, and additional species are noted. The next doubling is a 1 × 1 m quadrat, followed by 2 m^2, then 4, 8, 16, 32, 64 m^2, and so on. At each enlargement, the new species encountered are added. When plotting the cumulative number of species over the area sample, a species/area curve is obtained. In a homogenous vegetation segment, this usually takes the form of a steeply rising curve that levels off with each enlarged quadrat. The result is a curve of "diminishing returns" in terms of number of species encountered with still greater enlargement of area. The area over which the curve levels can be used as the "minimal area." The usual practice is to use a slightly larger area for the size of a relevé. In temperate grasslands, a 10 × 10 m quadrat is often sufficient to satisfy the minimal area requirement. In temperate forests, 20 × 20 m quadrats may give a satisfactory relevé size. In multispecies tropical rain forests, 1 ha (10,000 m^2) may not be large enough (see Fig. 4). But 1 ha is now often considered a practical standard for biodiversity research of tropical rain forests.

b. Recording Species in Layers

A relevé record should contain all plant species found in its boundaries. In forest communities, recording is best done in horizontal layers by vertically defined height strata. For example, one may distinguish two tree layers T1 as trees over 10 m tall, T2 as trees from 5 to 10 m tall (subcanopy trees), then a shrub layer from 2 to 5 m (including smaller trees), then a lower shrub layer from 1 to 2 m high, and further a herb layer from 0 to 1 m high, including smaller woody shrubs and trees (usually seedlings). The vertical stratification is an aid in recording. It also can document if canopy trees are regenerating in the same sample stand (or relevé).

c. Distributing Relevés in Predefined Strata

Relevés are best laid out in predefined strata. On mountain slopes, these may be predefined elevational strata, such as 200 m contour intervals or the like. In other areas, these are the tentative vegetation segments or communities identified on air photos in conjunction with field reconnaissance and check listing. We have called this process "entitation" (Mueller-Dombois and Ellenberg, 1974), which means defining of entities or map units from air photographs. Depending on the scale and homogeneity of the initial strata or entities, from 2 to 5 relevés are often used to characterize a stratum or entity. Also, depending on the size and complexity of plant biodiversity in an area, from 20 to 140 relevés can amount to a satisfactory biodiversity survey in terms of plant species and plant communities.

d. Estimating Species Quantities

The relevé method as described here is merely a formalized floristic checklist method, which reports the presence (and absence) of species. For classifying plant communities it is desirable to record how abundant or rare a species is in an area. This can be done simply by adding an abundance symbol to each species on the relevé record sheet. The most widely used rating system in European vegetation surveys is the Braun-Blanquet cover/abundance scale consisting of seven symbols as follows:

5	Any number of individuals, with cover more than 3/4 of the reference area (>75%)
4	Any number, with 1/2 to 3/4 cover (50–75%)
3	Any number, with 1/4 to 1/2 cover (25–50%)
2	Any number, with 1/20 to 1/4 cover (5–25%)
1	Numerous, but less than 1/20 cover, or scattered, with cover up to 1/20 (5%)
+	Pronounced cross) few, with small cover
r	Solitary, with small cover

This estimation scale is applied by walking diagonally through the relevé, several times if necessary. It is a crude scale and thus is often criticized, but it takes little time and tells a lot more about the plant composition of a relevé than a mere presence/absence record.

However, estimation methods are often not satisfactory in more complex vegetation. In forest vegetation it may be usefully applied only to the undergrowth species and only in relevés that are small enough to allow such estimation with confidence.

2. Quantitative Methods

In forest vegetation it is not possible to get a satisfactory estimate of the quantity of tree species by the cover/abundance scaling method as described earlier. In such more complex vegetation, at least the tree component needs to be measured to obtain a satisfactory abundance or cover estimate. The reason for this is primarily that relevé sizes are too small for obtaining an adequate density or cover estimate of trees.

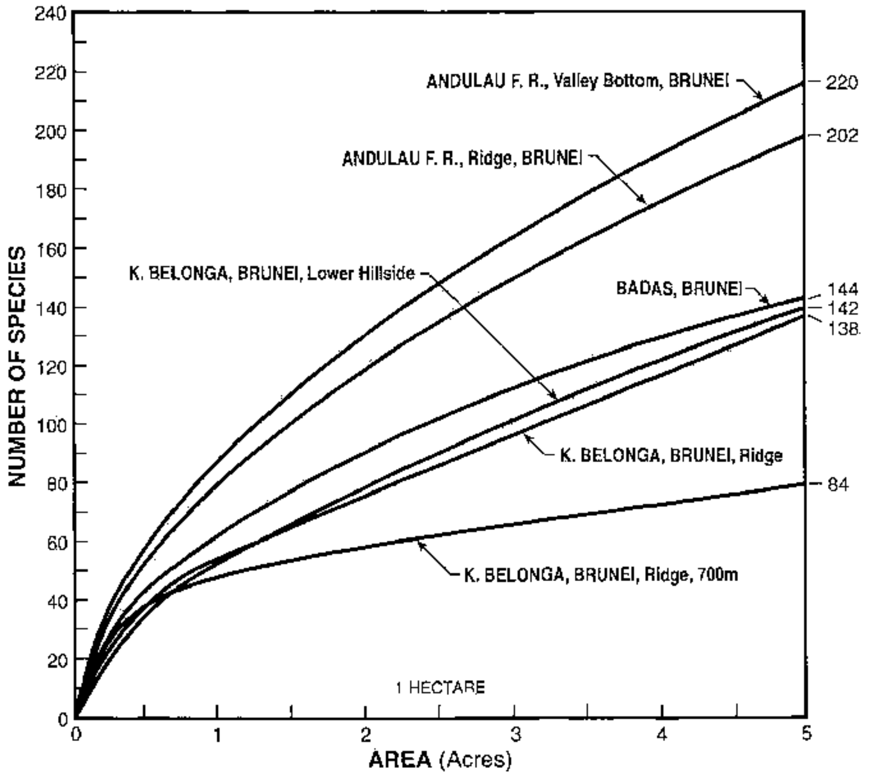


FIGURE 4 Species/area curves for sites with tropical rain forests in Brunei. Redrafted with modifications from Ashton, 1965.

a. The Count-Plot Method

To get an adequate estimate of tree density per unit area, such as a hectare, one should count trees in many small plots, such as 3 × 5 m. The small size of such count plots is merely a convenience for an accurate (100%) enumeration per plot. Larger plots may lead to accidental double counts or omissions, thereby introducing serious errors. Count plots are conveniently laid out along transects. They must be repeated until the cumulative mean number of trees does not fluctuate anymore radically with each additional count plot. This may be accomplished by counting at least 30 to 50 trees/species. It is obvious that rare tree species can hardly be adequately counted by this method. Rare tree species can only be reasonable assessed by searching an area of known size for such species, thereby attempting a 100% survey of their presence. In counting trees as a measure of tree density per unit area, it is necessary for plant biodiversity surveys to count trees by species and by measuring some tree-size parameters. Among tree-size parameters, usually the diameter at breast height (Dbh) is measured. From this, tree basal area is obtained by the area formula $r^2\pi$. However, the true basal area is the stem cover at ground level. Where this parameter is desired, one needs to measure also tree diameters at ground level. The Dbh measure is usually preferred in timber volume surveys. Tree volume is established by combining the basal area with a height measure and a tree form factor. More elaborate formulas can be developed for total tree biomass, often sought in productivity studies.

b. Distance Methods

An alternative to the count-plot method are the distance methods. They often lead more rapidly to an adequate count of trees, since they do not require the layout of plot boundaries. Instead, the average distance between trees is used to determine the mean area per tree and thereby the density per unit area. The mean area per tree is obtained by squaring the mean distance. This method works well under certain restrictions. Two methods are briefly outlined.

i. The Point-Centered Quarter Method A random starting point is established in the forest to be measured. Four distances are measured from this point to the nearest tree in four quarters, forming each a 90° exclusion angle over the point. The sampling point therefore is used like a rectangular cross, and the distance from the point to the nearest tree is measured in each quarter. When repeated at 20 sampling points, established at random or along a compass line, one obtains 80 distances. The mean distance squared then gives an estimate of the mean area per tree. The number of trees per acre (= 4 000 m^2) or hectare (10,000 m^2) is obtained by dividing such reference area by the mean area/tree. The point-centered quarter method is relatively simple to apply, and it compares well to the count-plot method. However, two restrictions apply: the trees should be randomly distributed, and no tree should be measured twice. The second requirement is usually easy to follow as it requires that the sampling points are far enough apart so that no tree is measured twice.

ii. The Wandering Quarter Method This second recommended distance method is often less affected by the distribution pattern of trees (random versus clumped) as it traverses for the same number of distance measures through a larger area. The method begins with establishing a random sampling point in the forest segment to be sampled. A 90° exclusion angle is established in a certain compass direction, and the nearest tree to the sampling point is measured. That nearest tree then becomes the next sampling point over which a 90° exclusion angle is established in the same compass direction. From this tree the distance is measured to the next tree in that exclusion angle. The procedure is continued for 20 distances in that same direction, moving from tree to tree and always using the same 90° exclusion angle. This results in a wandering movement or zigzag line depending on where the next tree occurs within the 90° exclusion angle. After 20 distances, the direction may be changed following a right angle for another 20 distances, then once more at a right angle in a parallel reverse direction to the first zigzag line, and finally, 20 distances in direction of the starting point. In this case, the sampling also includes 80 distances, but the area traversed is larger than in the point centered quarter method.

c. Measures Obtained in Count-Plot and Distance Methods

The two types of methods provide for the same three parameters desired often in quantitative vegetation surveys. These are frequency, density, and cover.

"Frequency" is simply the occurrence of a species in any number of plots in relation to the total number of plots in the sample. In the distance methods, frequency is the occurrence of a species at a sample point out of the total number of sample points. Frequency is a mixed measure of abundance and dispersion or distribution. It is a relative measure as it depends on the plot size used in sampling.

"Density" is the actual count of individuals per species in the total number of count plots. The sample plot area is then adjusted to a standard reference area, such as a hectare. In the distance methods, the number of individuals is obtained from squaring the mean distance of all trees and dividing this measure into the reference area. The number of individuals per species is then obtained from the proportion of species among the total number of individuals.

"Cover" as stem cover per species is obtained from totaling the individual basal area measurements in the sample of count plots or points and by relating this measure to the reference area. For examples, see Mueller-Dombois and Ellenberg (1974).

Another aspect relates to the tree-size stratification in count-plot and distance sampling. In addition to sampling mature trees/species, it is ecologically desirable in biodiversity inventories to also establish the density of seedlings and saplings. First, mature trees have to be defined. This is often done by setting height and/or diameter limits. "Mature trees," for example, may be all individuals 5 m and taller, thereby forming a broad size class. "Saplings" may be defined as individuals from 0.5 to 5 m tall and "seedlings" as those from 0.1 to 0.5 m tall. Smaller individuals may be considered as "germinants" (i.e. not yet well established or ephemeral individuals). Smaller subplots are often required for seedling counts because of their smaller size and frequently greater density. In the distance methods, usually only mature trees are measured because of the random pattern requirement for accuracy. In such cases, seedlings and saplings may be counted in subplots of appropriate sizes at predetermined sampling points.

Application of appropriate sampling methods requires some basic understanding of sample theory and field experience. When the sampling objective is clarified and the nature of the vegetation experienced, it is usually possible to prescribe an appropriate approach from a combination of established field analysis methods.

d. Size of Sample Plots

Figure 5 displays six sample plot sizes that have been used in proven studies. They are drawn to the same

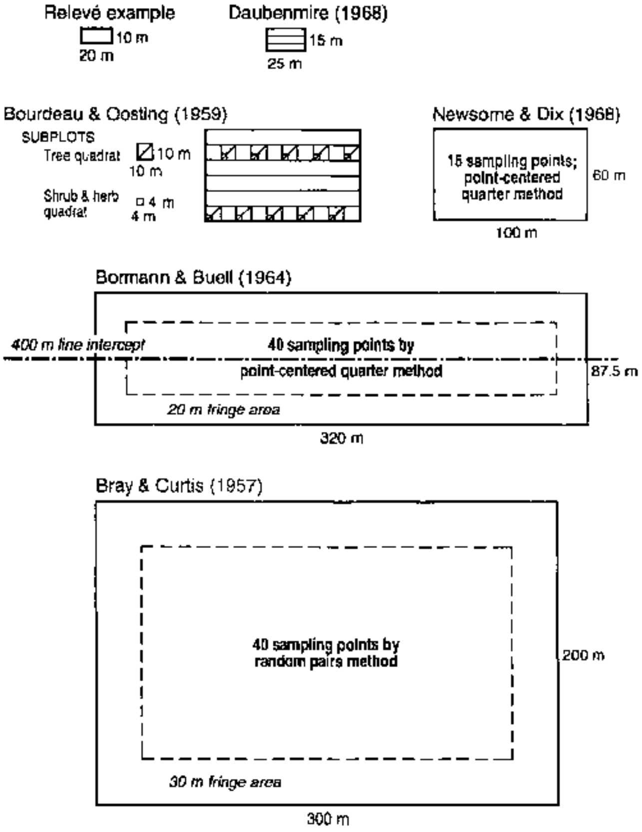


FIGURE 5 Comparative sizes of sample plots. From Mueller-Dombois and Ellenberg, 1974.

scale. Like the small, 20 × 10 m grassland relevé, all plots were placed in predetermined entities or tentative map units to obtain representative data. Daubenmire's (1968) preferred plot size of 15 × 25 m was applied in the mountainous interior forests of the Pacific Northwest. He preferred to assess the undergrowth vegetation within such plots along 25 m transects by determining frequency out of 50 systematically placed 0.1 m^2 frames. Percent cover in terms of shoot cover was also estimated similarly as with the Braun-Blanquet cover/abundance scale. But 50 placements only gave an area sample of 5 m^2, which likely did not fulfill the minimal area requirement as used in the relevé method. The larger plots in Fig. 5 refer to quantitative studies as discussed earlier. Their sizes are based on the principle that tree species need to be counted in numbers that give reliable density and basal area estimates. For obtaining a minimum tree count of 30 to 50 trees/species, larger plot areas are needed, since individual mature trees can occupy mean areas of 25 to 100 m^2 or larger.

The larger sample plots with quantification of tree counts obviously require a much greater time effort than a relevé with a species record supplied with a cover/abundance estimate. Thus, one has to be very clear about the purpose of the field effort. The relevé method is particularly useful in plant biodiversity studies aiming at classifying vegetation by species groups, while quantitative methods are essential if the objective is repeated monitoring of plant biodiversity in permanent plots.

C. Data Processing and Display Methods

1. Data Analysis: An Island Example

During the 1970s, we did a multidisciplinary study in Hawaii under the auspices of the International Biological Program (IBP). Our main objective was to study the biological organization in relatively undisturbed natural communities. We focused on two areas, an upper montane rain forest (the Kilauea Rain Forest) and the east-slope of Mauna Loa in Hawaii Volcanoes National Park (Fig. 6). In the Kilauea Rain Forest, which was dominated by three native keystone species, Acacia koa, Metrosideros polymorpha, and tree ferns (Cibotium glaucum), we established a large sample plot, 1000 × 800 m = 80 ha greater by 20 ha than the largest plot on Fig. 5. Our data sets were processed by species in life form and guild categories to get quickly from forest community structure to function by niche differentiation. This aspect is dealt with in part III of our synthesis volume (Mueller-Dombois, Bridges, and Carson, 1981), and will not be elaborated here.

On the east slope of Mauna Loa, we tried to answer a basic question asked by MacArthur (1972, p. 161), "Do different plant species change synchronously, or does each have independent distribution?" This question arose from a dispute over the nature of species distribution within and across neighboring vegetation zones or biomes. To clarify this question, we established 14 transect sampling sites (1–14 on Fig. 6) in predefined altitudinal intervals from 1100 to 3080 m. These sites were studied for 12 other organism groups besides plants. For the plant distribution study, we established 48 relevés throughout this mountain transect, approximately three relevés clustered around each of the 14 transect sites.

2. Two-Way Table and Dendograph Techniques

Following the field study, the 48 relevés were processed by the "two-way table technique." That required entering all relevés into a single "raw table," whereby the relevé number appears at the head of the table and all species names appear at the left side of the table. The

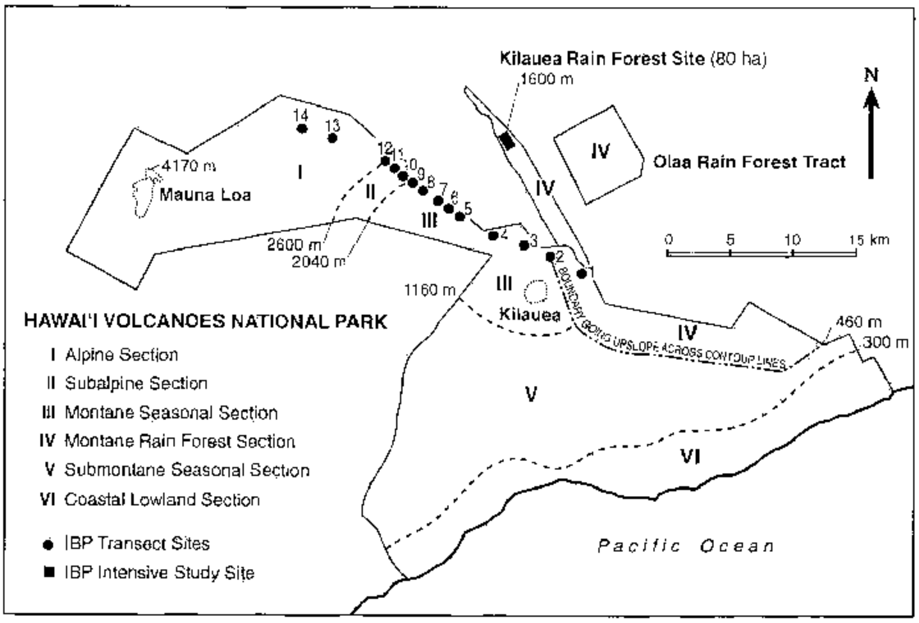


FIGURE 6 Map of Hawaii Volcanoes National Park with the 14 Mauna Loa transect sites and the 80 ha Kilauea forest site used for integrated biodiversity sampling during the Hawaii International Biological Program (IBP). From Mueller-Dombois, Bridges, and Carson, 1981.

TABLE II

Extract of a Final Two-Way Table

IBP Focal site no.	1			2			3			4			5			6			7			8			9			10			11			12			13						14					
	0	0	1	0	0	1	0	0	1	0	0	1	0	0	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	1	1	0	0	0	0	0	0	0	0	1	0	0	0	1	1	1
Relevé no.	0	5	4	4	5	4	1	5	4	1	2	4	2	9	0	9	0	0	8	0	1	2	1	2	8	2	2	7	2	3	7	3	3	1	1	7	4	6	6	6	6	3	6	6	7	3	4	4
	4	4	4	9	2	5	8	0	7	9	0	6	1	6	0	2	3	4	3	8	5	4	7	1	1	3	5	7	9	0	5	2	3	1	4	2	4	3	4	5	6	8	7	9	0	9	0	2
Species group 2																							*						*	*		*	*	*	*	*	*		*	*	*	*	*			*	*	
PELTER																											+			+		+	+	1	1	1	1	+	1	1	1	1	+			+	1	
ASP ADI																				R		+	R						+	+		+		1	1	1	R			R		R				+	R	
ASP TRI																							R		+	R				R		+		1	1	+	+		R	R	R						R	
AGR SAN							R																						R	R			1			+	1		1		+		+		+		+	1
																							*						*	*		*	*	*	*	*	*		*	*	*	*	*			*	*	
Species group 1					*		*	*																				*	*	*	*	*	*	*	*	*												
MAC GAH					+		1	1	1																			1	1	1	1	1	1	1	+	1										R	1	
COP ERN					3		+	+	1																		2	1	2	2	1	2	2	2	2	+												
LUZ HAW					1		1	2					+													+	1	1	1	1	+	1		R		+												
RAI CIL					R			1																				+	+	1		+	2	1	1	R												
COP MON																												1	1	2	1	2	2	2	2	2												
PLE THU								+		R																			+	+	1	+	+	1	+													
POL PEL				+	+		1	2																				+	+		1	+																
					*		*	*																				*	*	*	*	*	*	*	*	*												
Species group 3							*	*					*	*	*	*	*	*	*	*	*	*	*	*	*	*	*																					
HOL LAN					+		+	R		3	2	2	2	1	1	1	2	1	1	1	1			1	1	2	2																					
CAR WAH				+			R	+					2	+	1	1		1	1	2	2	2	1	1	1	1	1						1	+														
ACA KOA									2		2	2	5	4	3	3	3	3	4	4	3	5	4	3		3	2																					
STY TAM					2		3	3	3				2			1		3	2	2	2	1	2	3	3	2																						
PAN TEN													1	1	1	2	1	1	1	2	2	2	1	2	2	2																						
							*	*					*	*	*	*	*	*	*	*	*	*	*	*	*	*	*																					
Species group 4					*		*	*	*				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*												
DES AUS					1		+	2					2	1	1	3	1	1	3	2	2	4	2	2	2	2	2	1	1	1	1	1	1	1	1	1												
PTE AQU					1		2		2	2	3	2	2	2	2	1	2	1	1	+		1	+	+	1	1	2		1	1	1																	
DOD SAN					+		1	2	2							+		2	1	R	+							2	2	2		2	2	2	1	1												
					*		*	*	*				*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*	*												

Note: Rules used: 50% inside, 10% outside.

species magnitude, or score values, are entered for each species in the respective "relevé columns." When all data are entered in this way, one can see in the "species rows" which species are present (and with what score value) or absent in the "relevé columns." The process then begins of sorting species of similar distribution together by repositioning the "species rows," then also the "relevé columns" are resorted to bring those relevés together in the table that contain species of similar distribution. The "row and column sorting process" is reiterated until the table displays an optimal structure in the sense that one can interpret the clustering of species and relevé group in an ecologically meaningful way. This data processing, formerly done by hand, can now be done mathematically by using one of many multivariate analysis techniques and similarity indices.

An example of a two-way table extract of our 48 Mauna Loa relevés is given in Table II. For corroboration of species distribution and relevé clustering trends one can use various mathematical ordination methods or data classification methods. Ordination methods (not illustrated here) present the same data geometrically, while classification methods present the same data in the form of dendrographs (see Fig. 7). The three data processing and display techniques are corroborative analytical tools. Particularly the ordination and dendrograph techniques are rather abstract and difficult to interpret by the reader.

3. Improved Display Methods

It is more "reader friendly" to go a step further in displaying the results of a multivariate data analysis as shown in Fig. 8. This data display over the Mauna Loa transect with its 14 IBP transect sites clarifies that both multivariate analyses, the two-way table and the dendrograph techniques, gave closely similar results. Both data treatments resulted in the definition of seven altitudinal vegetation zones as named in the legend of Fig. 8.

The answer to MacArthur's (1972) question is given

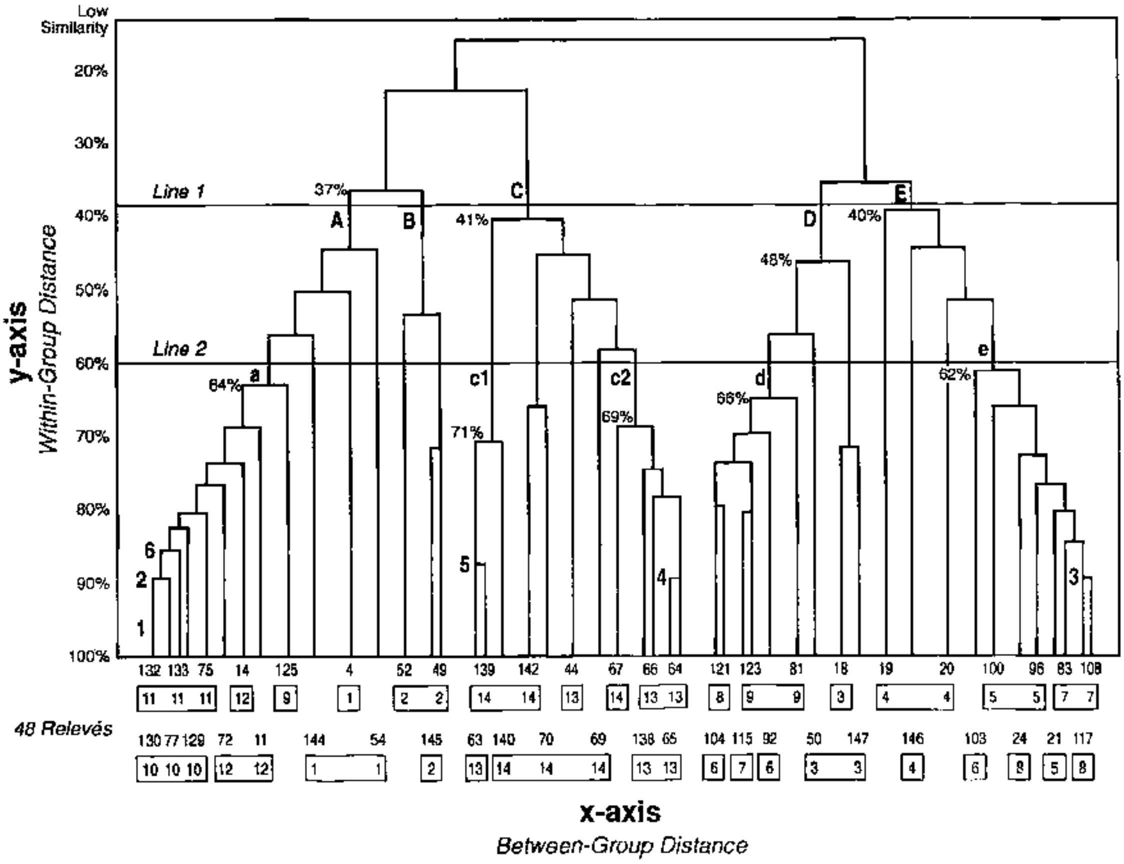
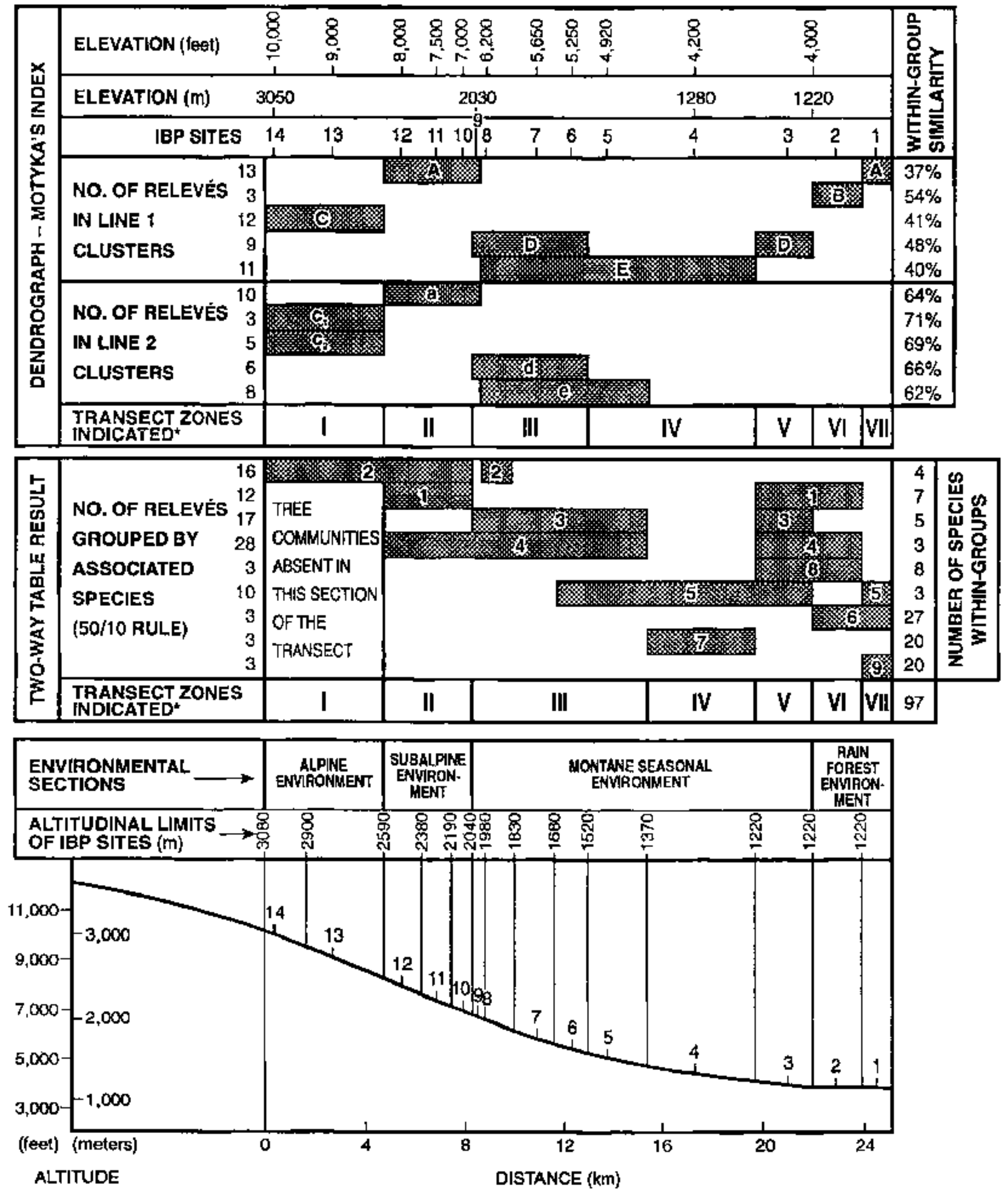


FIGURE 7 Dendrograph for cluster analysis based on 48 relevés sorted by the quantitative Sorensen index of similarity. Blocked out numbers are IBP transect sites 1–14. Numbers not blocked out are relevé numbers. From Mueller-Dombois, Bridges, and Carson, 1981.



*by at least two cluster or group distribution limits

FIGURE 8 Altitudinal vegetation zones on the east slope of Mauna Loa derived from computer processing of 48 relevés by dendrograph and two-way table techniques. Zone I = Sparse alpine heath scrub, II = Subalpine Metrosideros scrub forest, III = Mountain parkland with Acacia koa colonies, IV = Sapindus-Acacia koa savanna, V = Open Metrosideros dry forest, VI = Open Metrosideros rain forest, VII = Closed Metrosideros rain forest. From Mueller-Dombois, Bridges, and Carson, 1981.

in Fig. 9. It displays the distribution of 17 key species, which are primarily responsible for the pattern of the seven vegetation zones identified by multivariate techniques. Several species clearly change synchronously, while others have independent distributions along this island mountain slope. The answer to MacArthur's question is this: There is more than one pattern of plant species distribution along this geologically young tropical mountain slope. Some distributions coincide closely with one another; others do not. However, the

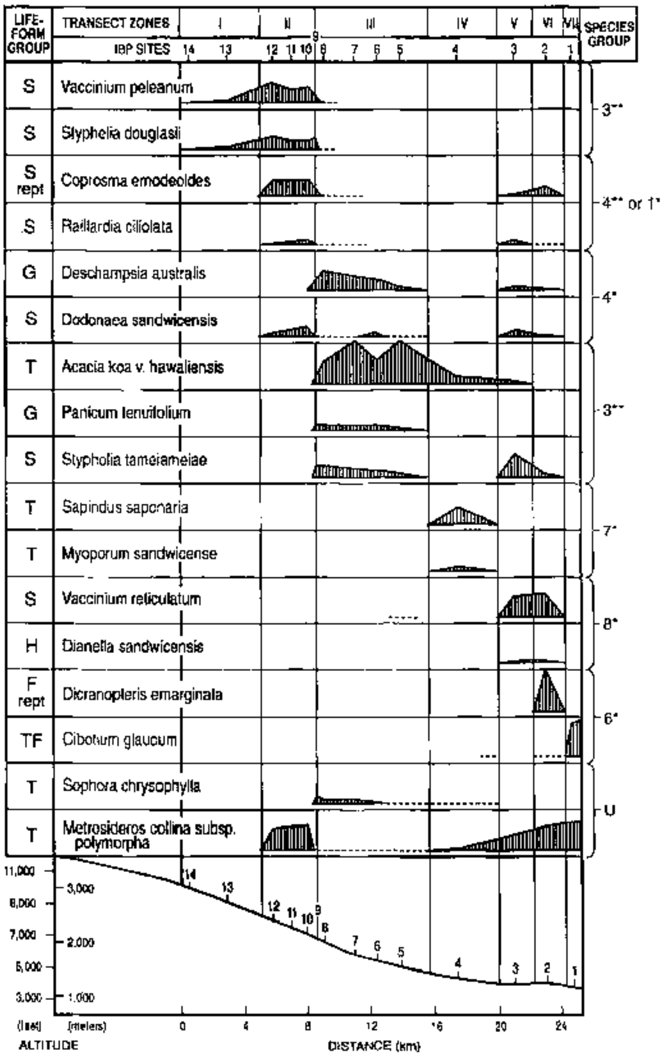


FIGURE 9 Distribution and abundance of native key species along the Mauna Loa transect. Species groups derived from two-way table analysis. Computer derived group numbers shown at right; u = ungrouped species. Symbols at left are, S = shrub, rept = reptate or creeping, G = grass, T = tree, H = herbaceous plant other than grass, F = fern, TF = tree fern. From Mueller-Dombois, Bridges, and Carson, 1981.

species distribution data give clear evidence of the existence of altitudinal vegetation zones as established from mathematically independent multivariate analysis techniques.

IV. AN EXPERIMENTAL SAMPLING DESIGN FOR ISLAND BIOGEOGRAPHY RESEARCH

Transects are generally the most efficient sampling designs in ecological field research. They can often be arranged to cut through a maximum of variation in representative habitats and biodiversity assemblages of an area to be studied, using the shortest distance. A still more important aspect is the use of transects to control environmental variation in its effect on biodiversity. In this way, transects can be used as "experimental sampling designs" by aligning them along changing environmental control factors to study their effects on biodiversity in environmental settings that are relatively uniform or constant. Sampling transects along a mountain slope, on which mean annual temperatures and rainfall change in predictable ways, are good examples. Another example is given in Fig. 2, in which the western Pacific atoll islands are aligned along mean annual rainfall gradients in both the northern and southern hemispheres.

A. Horizontal within Ecosystem Gradients

In addition to the atoll islands there are several other Pacific-wide biomes. A particularly important one is the upland rain forest on the volcanic islands of the tropical Pacific. This forest is important for two primary reasons: it is still extant on many volcanic high islands as indigenous forest harboring most of the endemic species, and it performs the function as watershed cover in the mountainous interiors of these islands. The volcanic high islands typically are small, isolated land fragments that protrude as mountainous terrain above the vast Pacific ocean surface. Their interior upland forests likewise are fragments that belong to the same larger ecosystem or biome. However, from archipelago to archipelago, these biome fragments are occupied by different sets of species due to their past biogeographic isolation. Connecting these biome fragments horizontally across the ocean by a system of transects provides for a within-biome research design. Here, the broad habitat features are kept uniform, while the indigenous biodiversity sets change from island to island. This truly is a research approach to biodiversity, which Pielou (1979) defined as "geoecology," the study of recurrence of similar communities in similar habitats, which are occupied by different sets of species. The term *geoecology* is an abbreviation of geographical ecology in the conceptual sense used by MacArthur (1972), who contributed substantially to the foundation of island biogeography.

B. Vertical between Ecosystem Gradients

A functioning watershed cover is an essential resource component in all Pacific high islands. From here, the fresh water flow begins to be regulated and then influences almost all lower lying island ecosystems, the freshwater wetlands, mangroves, estuaries, fish ponds, fringing reefs, the entire coastal zone, and agriculturally used lowland areas. The ecosystem services of the upland forest watersheds have not been studied in the Pacific islands, in spite of the fact that they have been part of the traditional land-use system in the Pacific high islands, in Hawaii known as the ahupua'a system. This vertically arranged land-use system was recognized in the Hawaiian Islands as comprising four integrated management zones, "wao la'au," upland forest or wilderness area, to be left alone, "kula," the more open foot hill region, "wao kanaka," the agricultural zone in the lowlands, and "kahakai," the coastal zone. This vertically arranged multiecosystem human support system proved to be an optimal land-management scheme for the indigenous islanders in the past. Vertically arranged transect sites, using the fresh water flow and hydrology as unifying parameter, may also prove to be a good organizing principle for interdisciplinary research focused on the function of biodiversity at the landscape level.

V. THE PABITRA INITIATIVE

The acronym PABITRA refers to the Pacific-Asia Biodiversity Transect outlined on Fig. 10. The transect system connects the high island archipelagoes with indigenous tropical rain forests still extant in their interior uplands. They range from east to west, from the paleotropical outlier islands, the Hawaiian archipelago and the Marquesas, to the biotically rich continental islands of southeast Asia. The PABITRA concept arose from the concluding chapter of a recent book (Mueller-Dombois and Fosberg, 1998) with the chapter heading "The Future of Island Vegetation." Simply put, the indigenous island vegetation only has a future if the scientific community and conservation managers redouble their efforts in learning to understand the function of island biodiversity and thereby take appropriate conservation measures. Another impetus came from a Japanese research initiative, called DIWPA, Diversitas in the Western Pacific and Asia. This program is the Pacific-Asia representative of the new worldwide program "DIVERSITAS," promoted by the International Union of Biological Sciences (IUBS) with support from UNESCO and ICSU (The International Council of Scientific Unions in Paris, France). The promotion of DIVERSITAS started effectively with a week-long forum titled "Biodiversity, Science and Development: Towards a New Partnership" in September 1994 (Younès, 1996). Since then, DIWPA has focused on a "Green Belt," a north-south transect in western Asia for monitoring

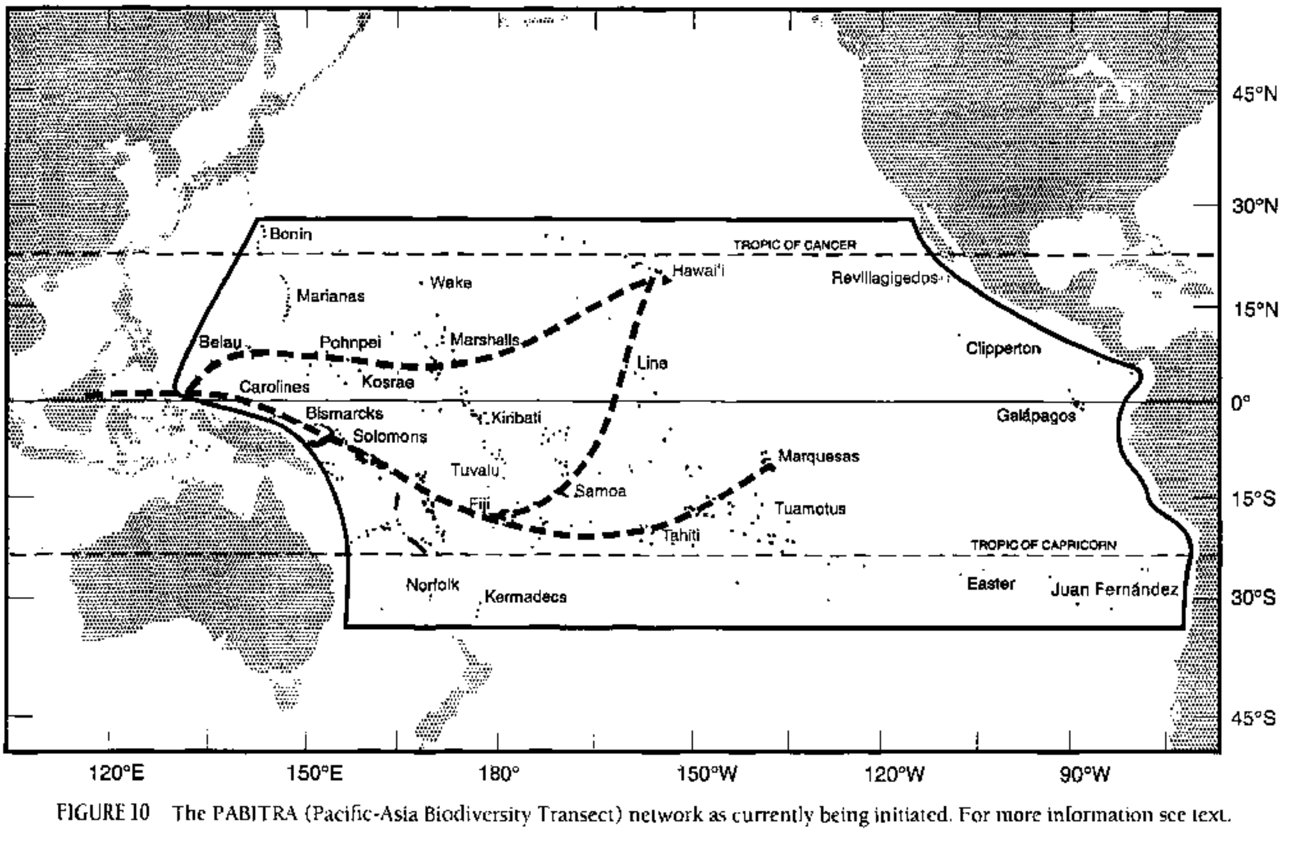


FIGURE 10 The PABITRA (Pacific-Asia Biodiversity Transect) network as currently being initiated. For more information see text.

biodiversity of forests and freshwater lake systems at 42 transect sites. A parallel running "Blue Belt" is also planned with focus on coastal zones and nearshore marine habitats. The PABITRA Initiative is considered a separate, but attached, "Pacific Island Branch of DIWPA," working cooperatively under the Ecology, Conservation, and Environmental Protection (ECEP) Division of the Task Force on Biodiversity in the Pacific Science Association, which is chaired by the writer.

See Also the Following Articles

BIOGEOGRAPHY, OVERVIEW • DISPERSAL BIOGEOGRAPHY • SPECIES AREA RELATIONSHIPS • SUCCESSION, PHENOMENON OF

Bibliography

Ashton, P. S. (1964). Ecological studies in mixed dipterocarp forests of Brunei State. Oxford Forestry Memoirs, no. 25.

Borman, F. H., and Buell, M. F. (1964). Old-age stand of hemlock-northern hardwood forest in central Vermont. Bull. *Torrey Bot. Club* 91(6), 451–465.

Bourdeau, P. F., and Oosting, H. J. (1959). The maritime live oak forest in North Carolina. *Ecology* 40(1), 148–152.

Bray, J. R., and Curtis, J. T. (1957). An ordination of the upland forest communities of southern Wisconsin. *Ecol. Monogr.* 27, 325–349.

Clements, F. E. (1916). Plant succession. An analysis of the development of vegetation. Carnegie Institute, Washington, D.C.

Cox, C. B., and Moore, P. D. (1993). *Biogeography: An Ecological and Evolutionary Approach*, 5th ed. Blackwell Scientific Publications, Oxford.

Crews, T. E., Kitayama, K., Fownes, J. H., Riley, R. H., Herbert, D. A., Mueller-Dombois, D., and Vitousek, P. M. (1995). Changes in soil phosphorus fractions and ecosystem dynamics across a long chronosequence in Hawaii. *Ecology* 76(5), 1403–1424.

Daubenmire, R. F. (1968). *Plant Communities: A Text Book of Plant Synecology*. Harper & Row, Evanston & London, New York.

Fox, R. L., de la Pena, R. S., Gavenda, R. T., Habte, M., Hue, N. V., Ikawa, H., Jones, R. C., Plucknet, D. L., Silva, G. A., and Soltanpour, P., (1991). Amelioration, revegetation, and subsequent soil formation in denuded bauxitic materials. *Allertonia* 6(2), 128–184.

Hengeveld, R. (1990). *Dynamic Biogeography*. Cambridge University Press, Cambridge.

Holdridge, L. R. (1967). Classification and characterization of tropical forest vegetation. *Recent Adv. Trop. Ecol.* II, 502–507.

Huggett, R. J. (1998). *Fundamentals of Biogeography*. Routledge, London, New York.

MacArthur, R. H. (1972). *Geographical Ecology: Patterns in the Distribution of Species*. Harper and Row, New York.

MacArthur, R. H., and Wilson, E. O. (1963). An equilibrium theory of insular zoogeography. *Evolution* 17, 373–387.

MacArthur, R. H., and Wilson, E. O. (1967) *The Theory of Island Biogeography*. Princeton University Press, Princeton, N.J.

Mueller-Dombois, D. (1986). Perspectives for an etiology of stand-level dieback. *Ann. Rev. Ecol. Syst.* 17, 221–243.

Mueller-Dombois, D., Bridges, K. W., and Carson, H. L. (Eds.) (1981). *Island Ecosystems: Biological Organization in Selected Hawaiian Communities*. U. S./IBP Synthesis Series 15. Hutchinson-Ross.

Mueller-Dombois, D., and Ellenberg H. (1974). *Aims and Methods of Vegetation Ecology*. John Wiley & Sons, New York.

Mueller-Dombois, D., and Fosberg, F. R. (1998). Vegetation of the tropical Pacific islands. *Ecol. Studies*, Vol. 132. Springer-Verlag, New York.

Newsome, R. D., and Dix, R. L. (1968). The forests of the Cypress Hills, Alberta and Saskatchewan, Canada. *Amer. Midland Naturalist* 80 (1), 118–185.

Pielou, E. C. (1979). *Biogeography*. John Wiley & Sons, New York.

Stoddart, D. R. (1992). Biogeography of the tropical Pacific. *Pac. Sci.* 46, 276–293.

Wagner, W. L., and Funk, V. A., (Eds.) (1995). *Hawaiian Biogeography: Evolution on a Hot Spot Archipelago*. Smithsonian Institution Press, Washington, D.C.

Walter, H., Harnickell, E., and Mueller-Dombois, D. (1975). Climate-diagram maps of the individual continents and the climatic regions of the earth. (Supplement to Vegetation Monographs). 36 pp. + 8 maps of world regions at 1:8 mill. And one world climate map at 1:30 mill. Springer-Verlag, Berlin, Heidelberg, New York.

Younés, T. (1996). Biodiversity science: issues and challenges. In *Biodiversity and the Dynamics of Ecosystems* (I. M. Turner, C. H. Diong, E. Shirley, L. Lim, and P. K. L. Ng, Eds.), pp. 1–12. International Network of DIVERSITAS in Western Pacific and Asia (DIWPA), Center for Ecological Research, Kyoto University, Shimosakamoto 4-1-23, Otsu 52001, Kyoto, Japan. DIWPA Series Vol. 1.

ISOPTERA

Takuya Abe[†] and Masahiko Higashi[†]
Center for Ecological Research, Kyoto University, Japan

GLOSSARY

alate An imago still possessing its wings.

caste A group of individuals in a colony that are both morphologically distinct and specialized in behavior.

imago The adult insects; the final developmental stage when insects possess wings.

larva An immature individual without an external sign of wing buds or soldier morphology.

neotenic A secondary reproductive with juvenile morphological characters. Neotenics derive from larvae, nymphs, pseudergates, or workers through at least one special moult.

nymph An immature individual on the developmental pathway to the imago and which possesses external wing buds.

presoldier A transitional morph that always precedes the soldiers; an unsclerotized individual whose head shows signs of soldier differentiation.

primary reproductives Dealate reproductives that founded a new colony after nuptial flight.

pseudergate (false worker) A temporarily nonreproductive individual serving the colony in nutrition, construction, or brood care, which results from a late, reversible deviation from the pathways to the imago and is characterized by reduced wing buds compared to nymphs of the same stage.

secondary reproductives Reproductives that differentiated in an established colony, whatever their origin and morphology. They may be supplementary reproductives if older reproductives are still present or replacement reproductives if not.

soldier An individual with a strongly sclerotized head showing defensive adaptations, such as enlarged mandibles, a stopperlike shape, or a frontal gland able to produce a defensive secretion.

queens and kings Females and males actively reproduce in a colony.

worker An individual resulting from an early, irreversible deviation from the pathway to the imago, and performing helper tasks. Workers are primarily characterized by the loss of the ability to proceed to the winged imago, but they need not be permanently sterile.

ISOPTERA (TERMITES) is a small order of insects containing about 2650 species. The oldest fossil is found in limestone deposits of 130 million years ago (Cretaceous). Termites consume cellulose, the most abundant organic matter on the earth, and are superabundant in tropical regions. Then termites are one of the most

---

[†] Deceased.

prosperous animals on the earth from the viewpoint of history length and abundance. This chapter presents the global diversification process of termites with special reference to their social evolution and digestive symbiosis with microorganisms.

I. INTRODUCTION

The Isoptera is a relatively small order of insects that consist of 31 orders, including about 2650 described species. The order name Isoptera originates from the fact that the imago (adult) has wings of equal size (*Isos* in Greek meaning the same and *ptera* meaning wings). The insects of Isoptera are commonly called termites in English. Due to their poor abdominal sclerotization and white appearance, they are often called "white ants" in many temperate regions. However, they are not phylogenetically closely related to ants, but to cockroaches and mantises. Termites are sometimes black, reddish brown, or yellow in body coloration in tropical regions.

Termites are widely distributed in tropical and subtropical regions, spreading from humid forests to savannas and even arid areas. Their biomass approaches 10 g wet weight/m^2. Only a few animal groups such as human beings, earthworms, herbivorous mammals in African savanna, and ants claim this order of biomass. The basis of this extraordinary prosperity of termites lies in their feeding habits. They consume cellulose, which is the most abundant organic matter on the earth. The ecological basis of their prosperity is their symbiosis with microorganisms and their highly developed social organization.

The symbiosis produces novel abilities that cannot be attained by either of the partners alone. In particular, the symbiosis between higher organisms and microorganisms often creates capabilities for exploiting new food resources such as dead plant material containing large amount of cellulose, which is a potential energy source but difficult for most animals including human beings to utilize. On the other hand, sociality enhances the efficiency of resource exploitation. Either of these or both together may cause a rapid growth and spatial expansion of the population, leading to diversification.

Animals in the tropical rain forests where the species diversity is the highest in the world are characterized by the dominance of social insects (termites, ants, social wasps, and social bees), and the dominance of symbiotic associations. Termites possess both of those attributes. Termites and ants occupy about one-third of total animal biomass in Amazonian tropical rain forests. The symbiosis and sociality in termites, together with their super abundance, brings them to play a keystone role of "super decomposer" of dead plant material in tropical terrestrial ecosystems. They consume and transform a large amount of nitrogen-poor dead plant material into nitrogen-rich termite body, which is in turn consumed by a great variety of animals ranging from ants and spiders to chimpanzees and human beings. Thus they form the basis for a large food web in the ecosystems, although they are well-known pests of agriculture and wooden buildings.

Furthermore, termites are prominent ecological engineers, modifying the soil properties by constructing huge mounds and long subterranean galleries and providing many animals and plants with hetrogenous habitats. It is also possible that they are an important player in global change scenarios because they have high biomass and emit a large amount of greenhouse gases such as methane as well as CO$_2$.

There are several landmarks in the process of global diversification of termites: (a) acquisition and loss of symbiotic flagellates, (b) social evolution, (c) change of life type from one-piece type to separate type, and (d) change of dominant feeding habits from wood feeding to soil feeding. This chapter explains an outline of the diversification of termites in relation to those events as well as their ecosystem functions.

A textbook of termites, *Biology of Termites*, was edited by Krishna and Weesner (1969, 1970), and monographs of termites were published by Grasse (1982, 1984, 1986). A new textbook, *Termites: Their Symbiosis, Sociality and Global Diversification*, was recently edited by Abe *et al.* (2000).

II. PROFILE OF TERMITES

A. Phylogeny and Fossil Records

The Isoptera, which includes 280 genera and more than 2650 described species, is a small order, considering that more than 0.9 millions species of insects have been described. Termites include seven families: Mastotermitidae, Hodotermitidae, Termopsidae, Kalotermitidae, Rhinotermitidae, Serritermitidae, and Termitidae. Principal genera of the world termite fauna and their distribution are given in Table I.

The phylogeny of termites is shown in Fig. 1. Three families, Mastotermitidae, Hodotermitidae, and Termopsidae, are the oldest living lineages. These three families together with Kalotermitidae, Rhinotermitidae, and Serritermitidae are called lower termites, which are

TABLE I

List of Families and Principal Genera of the World Termite Fauna, Together with Distributions of the Genera in Terms of Major Zoogeographic Regions

Division	Distribution
Family Mastotermitidae	
Mastotermes	Australia
Family Kalotermitidae	
Kalotermes	Worldwide
Cryptotermes	Worldwide, except for temperate Eurasia
Neotermes	Worldwide
Rugitermes	Polynesia, New World tropics
Glyptotermes	Worldwide, except for North Temperate Zone
Calcaritermes	New World
Family Termopsidae	
Archotermopsis, Hodotermopsis	Eurasia, Temperate Zone
Zootermopsis	North America
Stolotermes	Africa, Australia, New Zealand
Porotermes	Africa, Australia, New World (Chile only)
Family Hodotermitidae	
Hodotermes, Microhodotermes	Africa
Anacanthotermes	Asia, Africa
Family Rhinotermitidae	
Psammotermes	Eurasia, Africa
Coptotermes	Worldwide, mainly tropical
Heterotermes	Worldwide, mainly tropical
Reticulitermes	Eurasia, North America, temperate only
Prorhinotermes	Worldwide, tropical islands and shores
Termitogeton	Asia
Parrhinotermes	Asia, Australia
Schedorhinotermes	Asia, Africa, Australia
Rhinotermes	New World tropics
Family Serritermitidae	
Serritermes	New World tropics
Family Termitidae	
Subfamily Apicotermitinae	
Speculitermes	Asia, New World tropics
Anoplotermes	Africa, New World
Euhamitermes, Eurytermes, Indotermes, Protohamitermes	Asia
Apicotermes, Hoplognathotermes, Lepidotermes, Trichotermes	Old World tropics
Ahamitermes, Incolitermes	Australia
Microcerotermes	Worldwide except for North America
Amitermes	Worldwide
Drepanotermes	Australia
Gnathamitermes	North America, temperate
Subfamily Termitinae	
Basidentitermes, Crenetermes, Cubitermes, Euchilotermes, Fastigitermes, Foraminitermes, Megagnathotermes, Noditermes, Ophiotermes, Pericapritermes, Procubitermes, Thoracotermes, Promirotermes, Thoracotermes, Unguitermes,	Old World tropics
Cavitermes, Dentispicotermes, Neocapritermes, Orthognathotermes, Spicotermes, Spinitermes	New World tropics
Termes	Worldwide, in tropics
Angulitermes	Eurasia, Africa
Dicuspiditermes, Homallotermes, Microcapritermes Procapritermes	Asia
Capritermes	Madagascar
Subfamily macrotermitinae	
Acanthotermes, Allodontermes, Ancistrotermes, Protermes, Pseudocanthotermes, Sphaerotermes, Synacanthotermes	Africa
Macrotermes, Microtermes, Odontotermes	Asia, Africa
Subfamily Nasutitermitinae	
Eutermellus, Mimeutermes, Verrucositermes	Africa
Bulbitermes, Hirtitermes, Hospitalitermes, Lacessititermes	Asia, New Guinea
Grallatotermes, Trinervitermes	Asia, New Guinea, Africa
Armitermes; Constrictotermes, Convexitermes, Cornitermes, Curvitermes, Labiotermes, Paracornitermes, Procornitermes, Rhynchotermes, Subulitermes, Syntermes, Velocitermes	New World tropics
Tenuirostritermes	New World
Nasutitermes	Worldwide, in tropics

Based on Wilson, 1971.

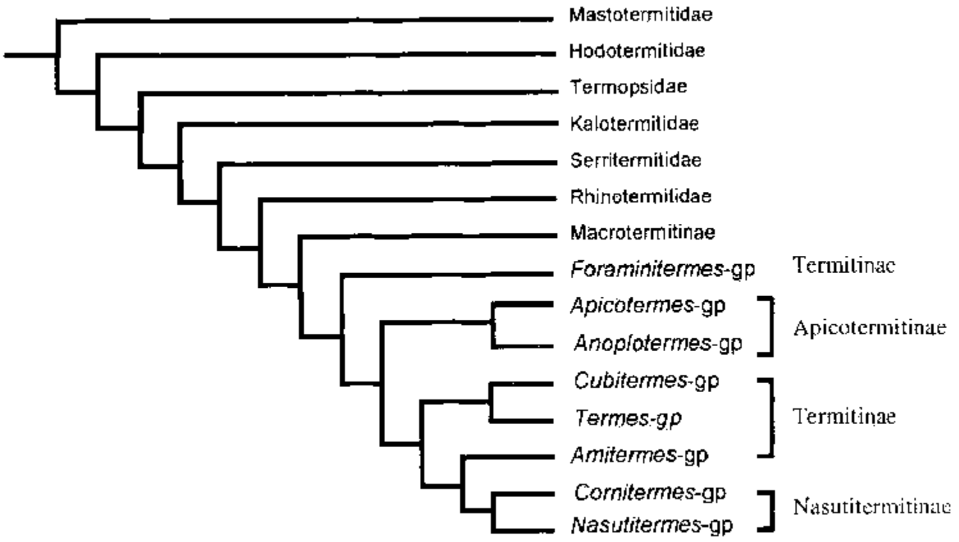


FIGURE 1 Phylogeny of termites. From Kambhampati and Eggleton, 2000.

characterized by the presence of symbiotic flagellates (Protozoa) in the gut. A remaining family, the Termitidae, is made up of the so-called higher termites, which are characterized by the absence of symbiotic flagellates, although some higher termites harbor symbiotic amoeba in the gut. The species diversity is the highest in Termitidae (the higher termites, ca. 1900 species), occupying about 70% of all species, while the Kalotermitidae have the highest species diversity (ca. 410 species) among the lower termites.

The fossil record of termites is fragmentary and has not been known from Africa. The number of fossil and living species of each family is shown in Table II. The oldest fossil is the worker of *Meiatermes bertrani* (Hodotermitidae) found in limestone deposits of Spain, dating to 130 million years ago (Cretaceous). A related species, *M. araripena*, was discovered in limestone deposits of Brazil, dating to 110 million years ago. The next oldest fossil species is *Valditermes brenanae* (Mastotermitidae) from England dating to 120 million years ago and *V. acutipennis* from Mongolia. Other Cretaceous species include *Cretatermes carpenteri* (Termopsidae), *Luteitermes prisca* (Hodotermitidae), and *Mastotermes sarthensis* (Mastotermitidae). Those Mesozoic termite fossils are found from Europe, Asia, and North and South America, indicating a broad Pangean distribution in both temperate and tropical habitats. A living termite, *Mastotermes darwiniensis*, whose distribution is confined to Australia, may well be called a living fossil.

Many fossil species of the Kalotermitidae and Rhinotermitidae, as well as Mastotermitidae, Hodotermitidae, and Termopsidae, have been found in Tertiary period, while only a few fossil species of the Termitidae (higher termites), which have the highest living species diversity, have been found in the Tertiary period.

Thus, the adaptive radiation of the higher termites seems to have occurred in the late Tertiary or Quarternary (Pleistocene) after the breakdown of Gondwana, yet they show a worldwide distribution. This

TABLE II

Numbers of Known Species of Termites

Groups	Cretaceous species	Tertiary species	Pleistocen species	Living species
"Lower" termites				
Mastotermitidae	3	19		1
Kalotermidae		33	0	411
Termopsidae	1	7	0	20
Hodotermitidae	3	14	2	15
Rhinotermitidae		16	1	305
Serritermitidae				1
"Higher" termites				
Termitidae				
Apicotermitinae				186
Termitinae		4		702
Macrotermitinae		3		332
Nasutitermitinae				659
Totals	7	96	3	2632

After Kambhampati and Eggleton, 2000; Thorne *et al.*, 2000.

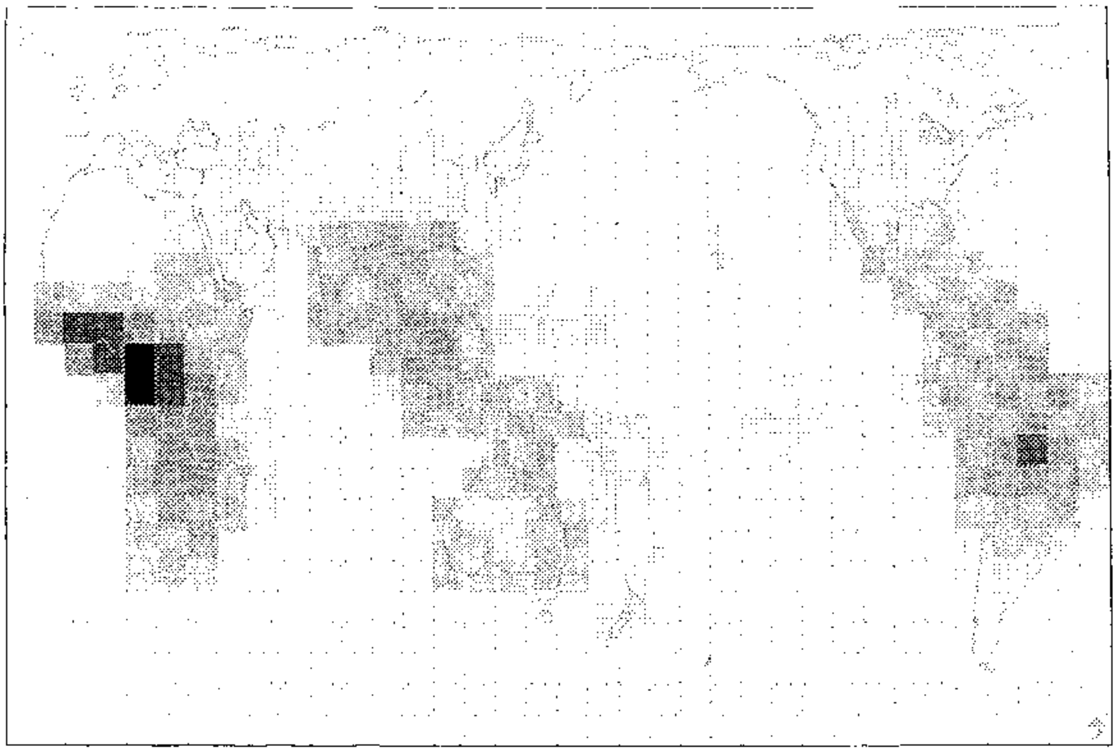


FIGURE 2 Global generic distribution of termites. The darker the square the higher the generic richness. Maximum generic richness is the southern Cameroon square. From Eggleton, 2000.

is an interesting problem remaining to be explained because the flight distance of termite alates is usually limited to some kilometers.

Based on the Cretaceous fossil record and living species diversity, two lower termite groups can be distinguished, primitive and modern. Primitive lower termites characterized by Createaceous species and low-living species diversity include Mastotermitidae, Hodotermitidae, and Termopsidae. The three families are the oldest living lineage as shown in Fig. 1. Modern lower termites characterized by moderate Tertiary and living species diversity include Kalotermitidae, Rhinotermitidae, and probably Serritermitidae.

B. Distribution

Global generic distribution of termites is shown in Fig. 2. Termites occur between the latitudes of 30–45N~40–45S and are mainly distributed in tropical regions, especially in western and central Africa, southeast Asia, northeastern Australia, and South America. The species diversity drops from 10 degrees north and south of the equator, but this gradient is not uniform on both sides of the equator (Fig. 3).

The three groups of termites—primitive lower, modern lower, and higher termites—show different distributions. Primitive lower termites show a fragmental or disjunctive distribution. The Mastotermitidae have only one species, *Mastotermes darwiniensis*, which is distributed in the savanna of northern Australia, although

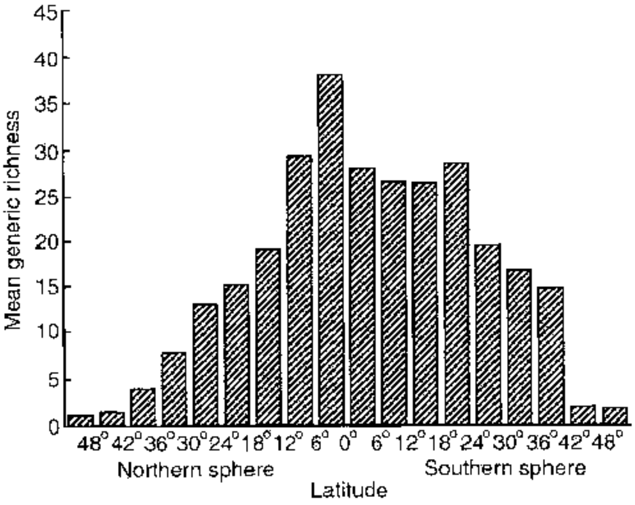


FIGURE 3 Latitudinal gradients of termite generic richness north and south of the equator. From Eggleton, 2000.

more than 20 species of Tertiary fossils are known from Europe, Australia, and North and South America. The Hodotermitidae including three genera (*Hodotermes*, *Microhodotermes*, and *Anacanthotermes*) and 15 species of grass feeders are distributed in the savanna and arid grassland of Africa and Asia although the family was widely distributed (Spain, Brazil, Europe, Japan, United States) in the Cretaceous and Tertiary periods. *Hodotermes* and *Microhodotermes* occur in Africa, while *Anacanthotermes* occurs in north African deserts, the Middle East, and southern India. The Termopsidae containing 20 living species show an amphitropical distribution: Three species of *Porotermes* are distributed in temperate regions of three southern continents (one species in each of Chile, Australia, and South Africa). Several species of *Stolotermes* are found in Australia, New Zealand, Tasmania, and South Africa, whereas the other three genera are distributed in the temperate and subtropical north hemisphere (*Archotermopsis* in Kashmir of India, *Zootermopsis* in North America, and *Hodotermopsis* in Japan, Vietnam, and China).

The modern lower termites except the Serritermitidae show a worldwide distribution. The Serritermitidae includes only one species, *Serritermes serrifer* in Brazil, and its phylogenetic position is under debate. The Kalotermitidae are distributed worldwide but tend to be distributed in coastal forests or islands. Among them, *Glyptotermes*, *Cryptotermes*, and *Neotermes* are pantropical, while *Kalotermes* shows a bipolar distribution. The Rhinotermitidae are widely distributed across tropical, subtropical, and temperate regions, and from wet to dry areas. For example, *Psammotermes* is one of the few insects persisting in deserts of Africa, the Middle East, and northwest India. *Coptotermes*, some species of which are serious pests of wooden buildings, are pantropical, partly because they are carried by humans. The distribution of *Heterotermes* and that of *Reticulitermes* are complimentary: *Heterotermes* is distributed in all tropical regions except for Africa, while *Reticulitermes* shows a Holarctic distribution. *Prorhinotermes* is mainly distributed on tropical oceanic islands.

The higher termites, including the four subfamilies of the Termitidae—Macrotermitinae, Apicotermitinae, Termitinae, and Nasutitermitinae—are characterized by central (tropical) distribution. The Macrotermitinae (fungus-growing termites) are distributed in the central areas of tropical Africa and southeast Asia. Three genera, (*Macrotermes*, *Odontotermes*, and *Microtermes*) are common. The Apicotermitinae include the *Apicotermes* and *Anoplotermes* groups. The *Apicotermes* group, most species of which are soil feeders, is distributed in tropical Africa, while the *Anoplotermes* group, which is a soldierless soil feeder, is distributed in tropical Africa, southeast Asia, and South America. The Termitinae are pantropical, including dominant genera of *Cubitermes*, *Termes*, *Amitermes*, and *Microcerotermes*. The Nasutitermitinae are also pantropical, including dominant genera of *Nasutitermes*, *Trinervitermes*, and *Cornitermes*.

It is interesting to note that the higher termites and some lower termites (the Termopsidae, the Kalotermitidae, and *Prorhinotermes* of the Rhinotermitidae) are complementary in geographic distribution: the former show a central distribution (occupying tropical mainland), while the latter show a marginal distribution (localized in temperate regions and tropical coastal areas and islands).

C. Termites as Eusocial Insects

An excellent introduction to eusocial insects was given by Wilson (1971). Eusociality is generally characterized by two generations of conspecific adults living together (that is, forming a group called a colony) and cooperation with each other to the extent that a strong reproductive skew is observed—that is, only one or a few adults in the group lay viable eggs (i.e., division of reproduction). The interior of the nest of a higher termite, *Amitermes hastatus*, is shown in Fig. 4. The primary queen (the largest individual) and king sit in the middle cell. Workers attend to the queen, to which they offer regurgitated food. Developmental pathways of the fungus growing termites are schematically shown in Fig. 5.

The eusociality of termites has evolved at least four times as sterile castes (once for the soldier caste and three times for the worker caste). Some wasps, some bees, all ants (Hymenoptera), and all termites (Isoptera) are typical eusocial insects, and sterile members have been found also in gall-forming aphids (Hemiptera) (soldiers), gall-forming thrips (Thysanoptera)(soldiers), an ambrosia beetle (Coleoptera) (workers), some mole-rats (workers), and a coral reef shrimp (soldiers), which are included in eusocial animals.

Termites have similar social organization to the Hymenoptera but differ from them in several important aspects (Table III). Termites are characterized by the diversity of soldiers, while the Hymenoptera are characterized by the diversity of workers. Furthermore, in termites, (a) females and males are both diploid (whereas females are diploid and males are haploid in the Hymenoptera), (b) females (queens) and males (kings) both take care of brood and male remains with female through life (whereas males are produced just before the mating season and die soon after mating in

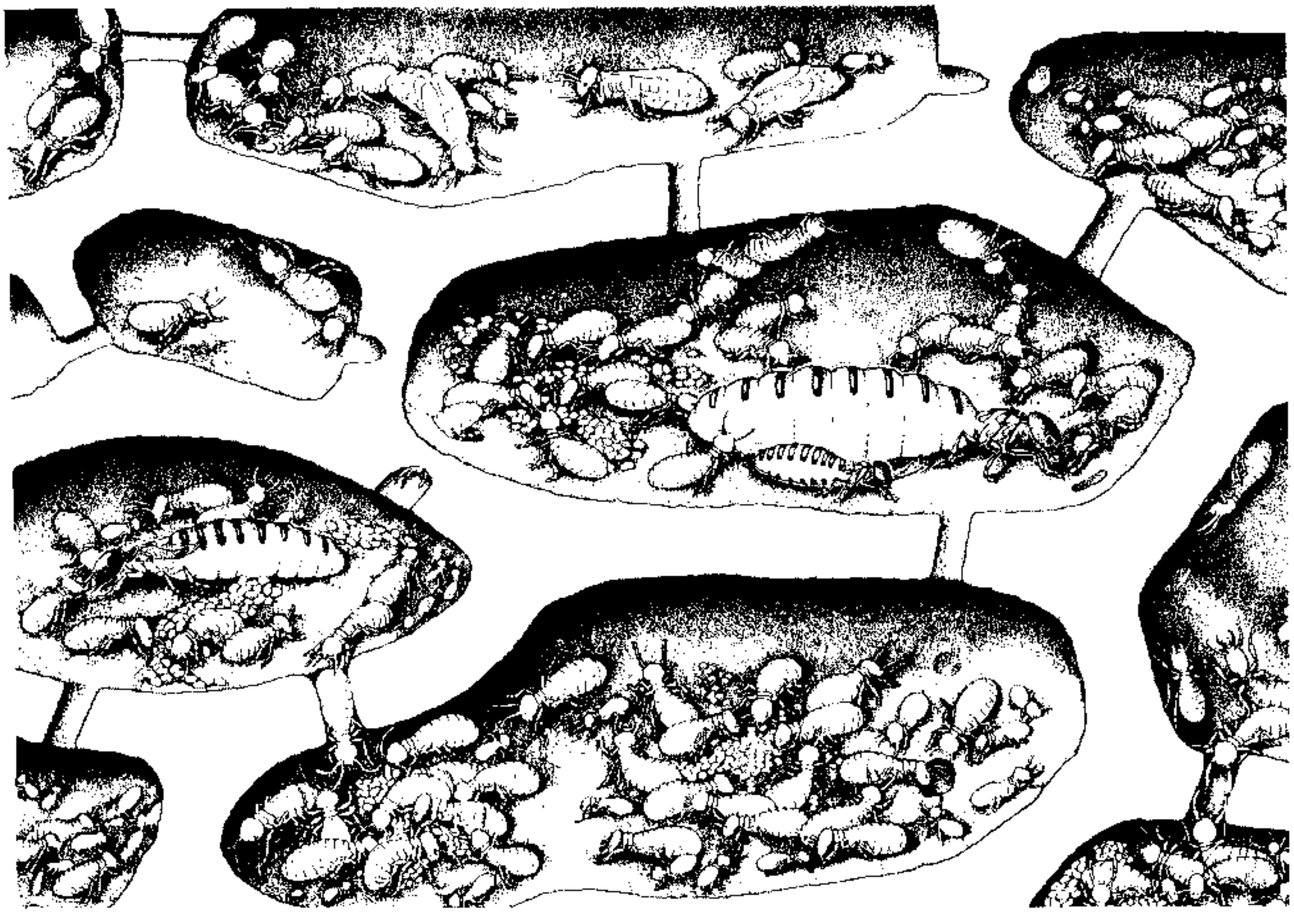


FIGURE 4 The interior of a nest of *Amitermes hastatus* of South Africa. The primary queen and male sit side by side in the middle cell. To the lower left is a secondary queen, which is also functional in this case. In the chamber at top center are reproductive nymphs, characterized by their partially developed wings. Workers attend the queens and are especially attracted to the heads, to which they offer regurgitated food at frequent intervals. Other workers care for the numerous eggs. A soldier and presoldier (nymphal soldier stage) are seen in the lower right chamber, while worker larvae in various stages of development are found scattered through most of the chambers. From Wilson, 1971.

the Hymenoptera), (c) workers and soldiers are either male or female or both (whereas they are all females in the Hymenoptera), and (d) termites are hemimetabolous (whereas they are holometabolous in the Hymenoptera).

D. Food and Predators

Animals are divided into two large groups in relation to their feeding habits: cytoplasm consumers and cell wall consumers. Most termites are cell wall consumers, whereas the Hymenoptera, including social insects such as bees, wasps, and ants, are cytoplasm consumers. The cell wall of plants consisting of cellulose, hemicellulose, and lignin is very poor in nitrogen but tends to be abundant and clumped, while cytoplasm is rich in nitrogen but tends to be sparsely distributed. Cellulose, a major component of plant cell wall, is decomposed to glucose, an important source of energy and body construction for animals. However, cellulose is difficult for most animals to decompose, especially when it is combined with lignin, which is extremely resistant to chemical degradation.

Termites as a whole feed on live and dead vegetation, lichens, algae, fungi, dung, soil rich in organic matter (so-called humus), other individuals of the same colony (cannibalism), organic-rich portion of termite nests, and skins of vertebrate corpses (Wood, 1978). Most termites in natural ecosystems do not attack living plants, although *Microcerotermes* sometimes attacks native living trees in Malaysia and some termites consume living native grasses in savannas and grasslands.

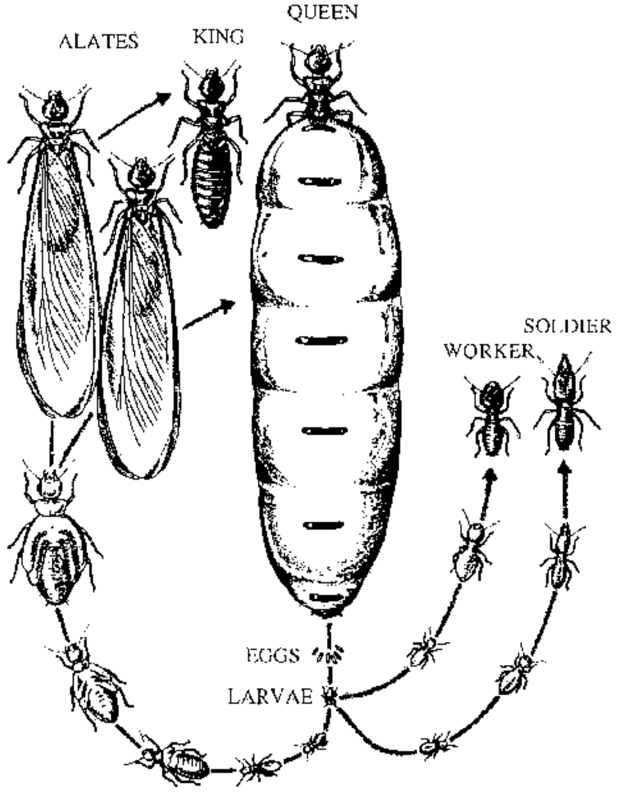


FIGURE 5 Schematic illustration of the developmental pathway of higher termites. From Behnke, 1977.

Lower termites eat dead wood except Hodotermitidae, which are grass feeders, while higher termites are rich in feeding habits. Some higher termites take specialized food. Some genera of the Nasutitermitinae feed on lichens on tree trunks: *Hospitalitermes* in southeast Asia, *Constrictotermes* in South America, and *Grallatotermes* in east Africa, India, the Philippines, and New Guinea. *Ahamitermes* and *Incolitermes* in Australia and *Ophiotermes* and *Tuberculitermes* in Africa are obligate inquilines in the nest of other termite species and feed exclusively on the carton of nest material of their hosts.

Soil feeders, which are recognized in the Termitinae, Apicotermitinae, and Nasutitermitinae of higher termites, include 1100 species, which is approximately 45% of the described termite species. The development of soil feeding is a landmark in the evolution of termites, considering that it is rare among animals. The real food source for soil feeders is unknown.

Termites are included in the diets of a wide range of animals. Ants are the most dominant predators. They are either obligatory or opportunistic predators. Some ponerine ants, such as *Termitopone commutata* in South America, *Leptogenys chinensis* in Sri Lanka, and *Megaponera foetens* and some subterranean *Dorylus* spp. in Africa, are typical obligate predators. The ant with lethal effects on a termite colony changes with the colony's development. For a colony of the fungus growing termite, *Macrotermes michaelseni* in Kenya, it is opportunistic predatory ants (individual foraging ants such as *Pachycondyla* and *Ophthalmopone* and ants with recruitment system such as *Camponotus* and *Pheidole*) at their early stage, ants with marked recruitment by trail pheromone (e.g., *Megaponera foetens*) at the middle stage, and subterranean group foraging doryline ants (e.g., *Dorylus juvenculus*) at the late stage.

Mammals that specialize in digging out termite nests are pangolins in Africa and India and aardvarks in tropical Africa. Other vertebrates specializing in predation on termite brood centers include armadillo in Texas of the United States, anteaters in South America, sloth bear (*Melurus ursinius*) in India, and the numbat and echidna in Australia. Wild chimpanzees in Africa make and use tools for "fishing" termites (Fig. 6).

E. Nests and Life Types

Termites are ecosystem engineers. They construct nests of various shapes and sizes, sometimes reaching to 5 m in height (Fig. 7). The diversity of nests is related to the diversity of social evolution, colony size, and feeding habits as well as the establishment of a microclimate suitable for termites to defend their colony against predators. The inner parts of carton nests of some species contain a large amount of wood fibers, which are eaten by the builders or utilized by inquiline termites. The most remarkable example is fungus gardens of Macrotermitinae.

Nest material has two origins: exogenous material such as particles of soil and wood transported by colony members and fecal. The cohesion of the nest construction is achieved by saliva moistening of exogenous material. Termites usually use both of fecal and exogenous materials for nest construction except for the Macrotermitinae, which do not use fecal material. Soil feeders and arboreal nest builders mainly use fecal material. Soil material is important in most epigeous nests (Abe, 1987).

Three life types are distinguishable in relation to nest and feeding sites: (a) a one-piece type, which nests in a piece of wood and consumes only that piece, (b) an intermediate type, which nests in a piece wood and consumes it as food but also constructs galleries to consume other wood pieces, and (c) a separate type, which nests in various sites such as on living tree trunks, in soil, on the ground surface, and so on, constructing

TABLE III

Basic Similarities and Differences in Social Biology between Termites and Higher Social Hymenoptera (wasps, ants, bees)

Similarities	Differences	
	Termites	Eusocial Hymenoptera
1. The castes are similar in number and kind, especially between termites and ants	1. Caste determination in the lower termites is based primarily on phermones; in some of the higher termites it involves sex, but the other factors remain unidentified	1. Caste determination is based primarily on nutrition, although pheromones play a role in some cases
2. Trophallaxis (exchange of liquid food) occurs and is an important mechanism in social regulation	2. Most species possess soldiers	2. A few species possess soldiers
3. Chemical trails are used in recruitment as in the ants, and the behavior of trail laying and following is closely similar	3. The worker castes consist of both females and males	3. The worker castes consist of females only
4. Inhibitory caste pheromones exist, similar in action to those found in honeybees and ants	4. Larvae and nymphs contribute to colony labor, at least in later instars	4. The immature stages (larvae and pupae) are helpless and almost never contribute to colony labor
5. Grooming between individuals occurs frequently and functions at least partially in the transmission of pheromones	5. There are no dominance hierarchies among individuals in the same colonies	5. Dominance hierarchies are commonplace, but not universal
6. Nest odor and territoriality are of general occurrence	6. Social parasitism between species is almost wholly absent	6. Social parasitism between species is common and widespread
7. Nest structure is of comparable complexity and, in a few members of the termitidae (e.g., *Apicotermes*, *Macrotermes*), of considerably greater complexity; regulation of temperature and humidity within the nest operates at about the same level of precision	7. Exchange of liquid anal food occurs universally in the lower termites, and trophic eggs are unknown	7. Anal trophallaxis is rare, but trophic eggs are exchanged in many species of bees and ants
8. Cannibalism is widespread in both groups (but not universal, at least not in the Hymenoptera)	8. The primary reproductive male (the "king") stays with the queen after the nuptial flight, helps her construct the first nest, and fertilizes her intermittently as the colony develops; fertilization does not occur during the nuptial flight	8. The male fertilizes the queen during the nuptial flight and dies soon afterward without helping the queen in nest construction

Modified from Wilson, 1971.

galleries to consume various dead plant material outside of the nest.

The one-piece type is supposed to be the most primitive. An evolutionary trend of life type is a progressive separation between food and nest. The first step is the development of a subterranean gallery system, which allows the colonization of new wood pieces (intermediate type), and the second step is the distinction of the nest from food (separate type). This change of life types seems to have occurred independently at least three times (Mastotermitidae, Hodotermitidae, and Rhinotermitidae) in the evolution of termites. The one-piece type is observed in primitive lower termites (Termopsidae) and modern lower termites (Kalotermitidae and *Prorhinotermes* of Rhinotermitidae). The Termopsidae usually nest in large damp wood and are called damp wood termites, while the Kalotermitidae usually nest in dry wood of living and fallen trees and are called dry wood termites. *Prorhinotermes* of the Rhinotermitidae nest in ordinary wood. In the one-piece type, the size, growth, and longevity of the colony are constrained by the size of a wood piece.

The intermediate type is observed in primitive lower termites (Mastotermitidae) and the modern lower termites (most Rhinotermitidae). The separate type is observed in primitive lower termites (Hodotermitidae),

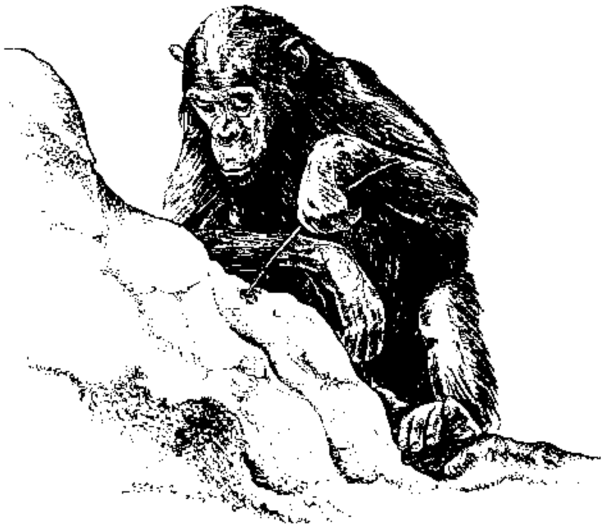


FIGURE 6 A chimpanzee fishing termites with a stick in a fungus-growing termite mound. From Behnke, 1977.

modern lower termites (a few Rhinotermitidae), and the higher termites (Termitidae). Separate type nests are classified into three categories: (a) subterranean nests constructed below ground, (b) epigeous nests protruding above soil surface, and (c) arboreal nests built on a tree trunk or a tree branch but always linked with soil by covered galleries.

III. SYMBIOSIS

A. Digestive Tube

Termite symbiosis is an obligate nutritional mutualism between the termite and microorganisms in the gut and nest of termites, although symbiosis, in the general sense, means the living together of two or more organisms that are not closely related in phylogeny, without the implication of beneficial exchanges.

The digestive organ of lower termites is schematically shown in Fig. 8. The digestive tube consists of foregut, midgut tubular (site for secretion of digestive enzymes and for absorption of soluble nutrients), and hindgut (voluminous site for digestion and for absorption of nutrients). The swollen portion of the hindgut is called the paunch. Malpighian tubules, which transport urine and urinary metabolites for excretion, empty at or near the junction of mid- and hindgut. Midgut and salivary glands produce the enzyme for decomposing cellulose into glucose. In the hindgut, cellulose is decomposed into acetic acid by symbiotic flagellates.

The microenvironment becomes more anoxic from foregut to hindgut. The hindgut used to be considered an anaerobic fermentation chamber, but recently this has been found to be false. The hindgut consists of an anoxic lumen surrounded by a microoxic periphery (Fig. 9). This is consistent with the occurrence of both anaerobic and O_2-dependent microbial metabolism in

FIGURE 7 Nests of *Nasutitermes triodiae* in Australia. See also color insert, Volume 1.

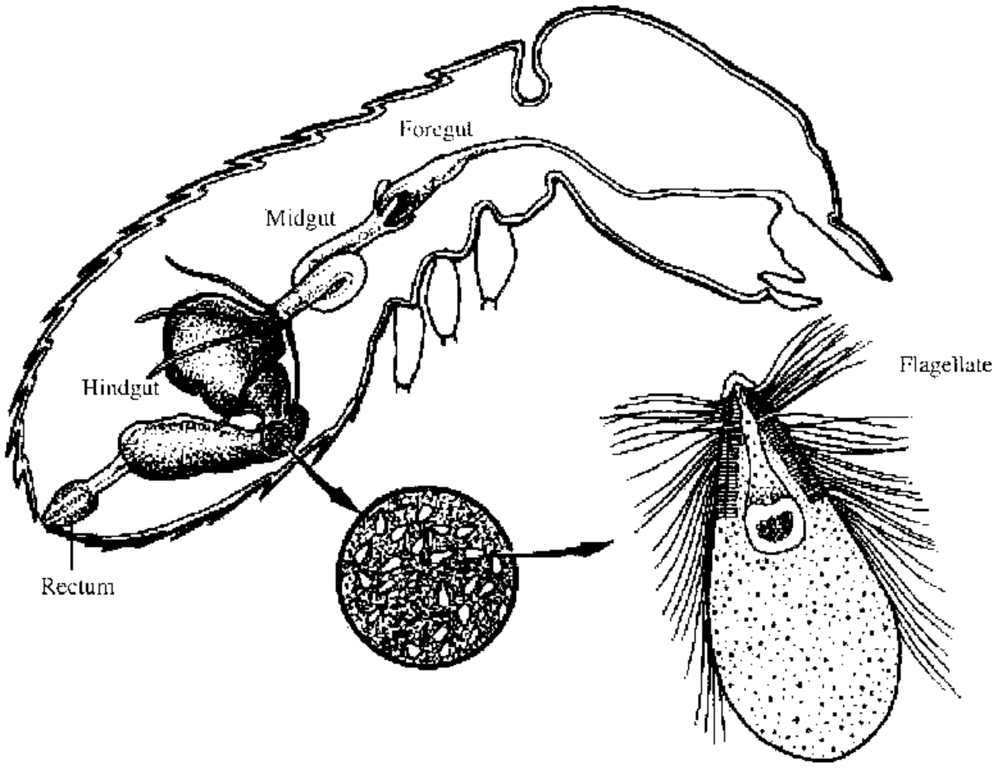


FIGURE 8 Schematic illustration of digestive system of a lower termite. From Behnke, 1977.

the hindgut. Bacteria on or near the gut wall constitute an oxygen sink, consuming the inwardly diffusing O_2 and thereby creating anoxic conditions favorable for fermentative production of acetate, a major energy source for termites. The hindgut of termites typically has a low redox potential (from −50 to −270 mV) and a pH around neutrality (6.2 to 7.6), although some portions of the hindgut of soil-feeding termites have a pH as high as 11.

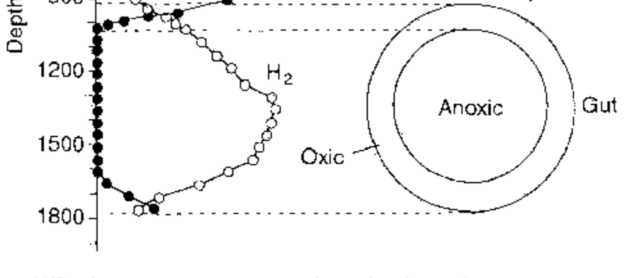


FIGURE 9 Microenvironment of the hindgut of a termite. From Breznak, 2000.

B. Two Types of Symbioses

Woody tissue, a major food of termites, contains only 0.03 to 0.1% nitrogen and their C/N (carbon/nitrogen) ratio is 350 to 500, whereas termite tissues contain 5 to 14% nitrogen and their C/N is 4 to 12. Therefore termites have to solve the "carbon-nitrogen balance" problem in addition to "cellulose decomposition problem" (Higashi *et al.*, 1992). Termites use free-living and symbiotic fungi, gut bacteria, and protozoa for assistance in decomposing cellulose and metabolizing and conserving nitrogenous compounds.

Thus, two symbioses are distinguished in termites: "cellulose decomposition symbiosis" and "C-N balance symbiosis" (Fig. 10). All the species of termites are, though to different extents, able to produce their own cellulase. This seems reasonable because they have no way to gain glucose, which is used not only for energy production but also for biological synthesis, other than

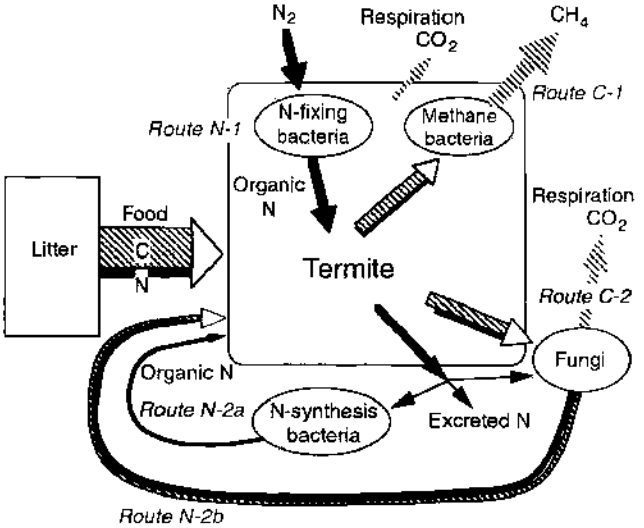


FIGURE 10 A summary (superimposition) of the possible mechanisms for C-N balance, with the responsible symbionts. Black symbols, nitrogen (N); hatched symbols, carbon (C). From Higashi and Abe, 1997.

digesting cellulose for themselves. The lower termites, which harbor symbiotic flagellates, need cellulose decomposition symbiosis to increase the efficiency of cellulose decomposition.

On the other hand, C-N balance symbiosis is necessary for all termites, because they are not able to solve the problem alone. Two mechanisms are possible for C-N balance: adding N to their food, referred to in Fig. 10 as route N-1, and selectively outputting C, referred to as routes C-l and C-2. Termites obtain N either through their food or by fixing atmospheric N. Termites tend to feed selectively on food with above average N context. Some species feed on plant tissue only when it is partially decomposed, and by this means, take advantage of the reduced C/N ratio that results from fungal and microbial decomposition. As selectively eliminating carbon is a "wood-consuming" manner, one-piece type termites, which stay in their wood nest and consume it, are expected to develop mechanisms to add N, such as symbiosis with nitrogen-fixing bacteria. On the other hand, separate type termites that are free from the limit in resource utilization are expected to make a full choice among possible mechanisms for C-N balance, including association with methanogenic bacteria. Among the separate type termites, the fungus-growing habit brings additional mechanisms for C-N balance; termites both add N through local N-recycling (route N-2b) and selectively eliminate C through fungal respiration (route C-2).

C. Cellulose Decomposition Symbiosis

The phylogenetic trend of symbiotic microorganisms in termites and cockroaches is shown in Fig. 11. Lower termites harbor flagellates in their hindgut. More than 400 species of flagellates of three orders, Oxymonadida, Trichomonadida, and Hypermastigida, have been described from 205 species of lower termites. Each species of lower termites usually harbors more than one species of flagellates. The class Parabasalea, to which Trichomonadida and Hypermastigida belong, is one of the oldest groups among eukaryotes, lacking mitochondria and peroxisomes. Members of this class use hydrogenosomes for energy production and, like the prokaryotes, have a 70S ribosome.

All colony members of lower termites except eggs, newly hatched larvae, and individuals just after molting harbor flagellates. They are transmitted through proctodeal feeding (i.e., transmission of gut contents from anus of a donor to mouth of a receptor). The flagellates of the mother colony are transmitted to offspring colonies by alates. The paunch of the worker is filled with flagellates and bacteria; they occupy 61% by weight of the hindgut contents of *Reticulitermes flavipes* (Rhinotermitidae). The population of flagellates in a gut of *Reticulitermes speratus* often reaches more than 105 individuals.

The association between lower termites and flagel-

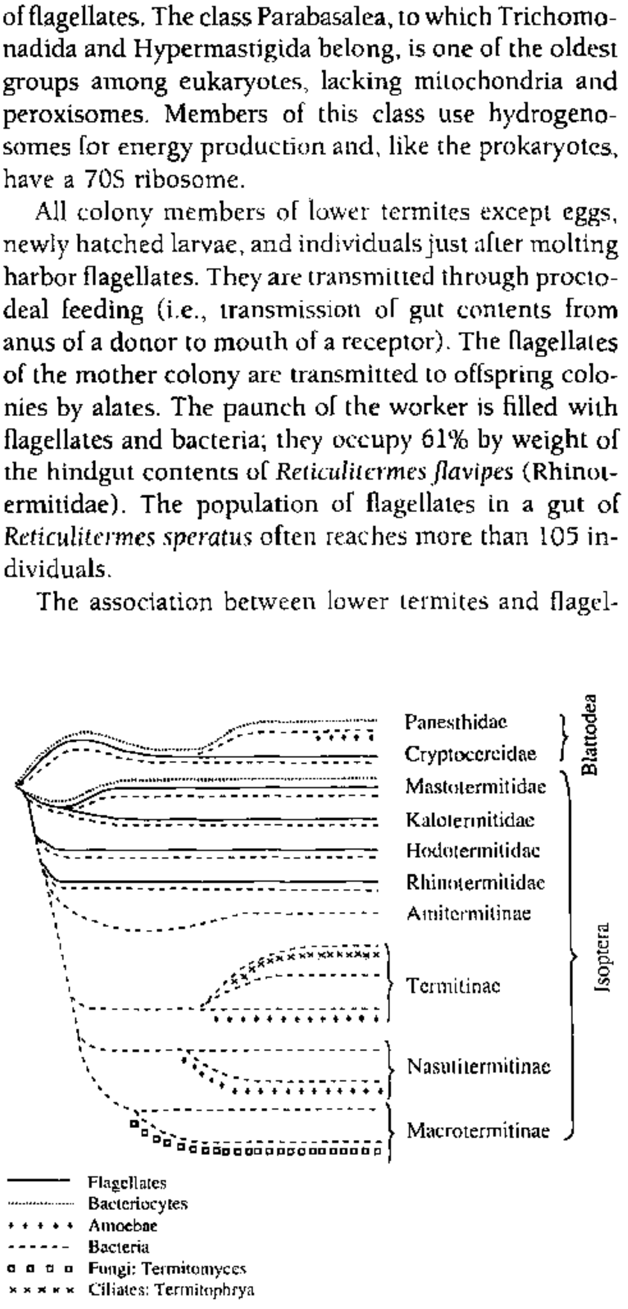


FIGURE 11 Phylogenetic diagram of the termites, in relation to the evolution of symbiosis with protozoans, bacteria, and fungi. Based on Grasse and Noirot, 1959, with some alterations.

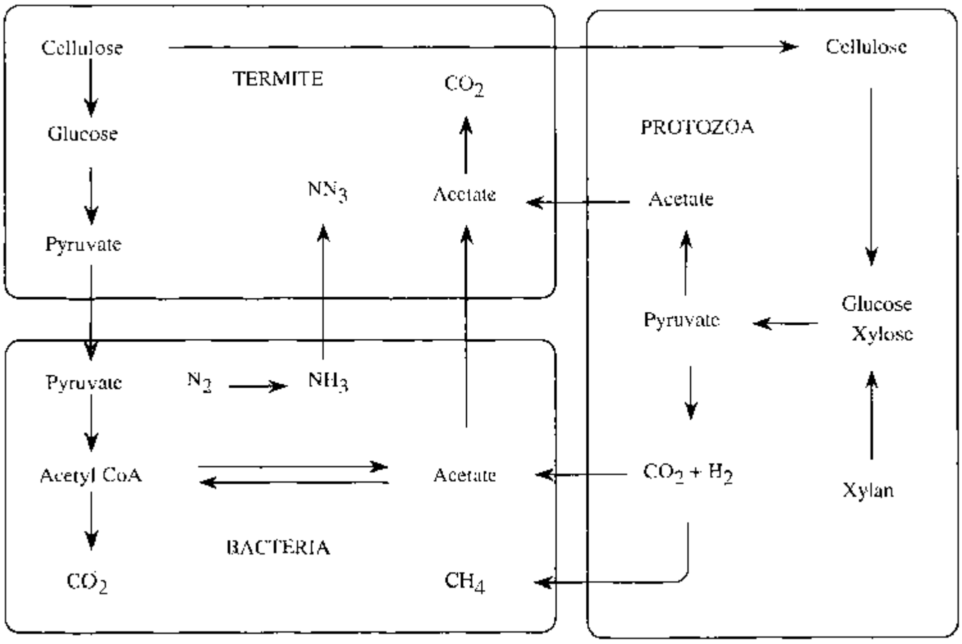


FIGURE 12 Biochemical overview of the major processes involved in the symbiosis between termites and their gut microbiota. From Slaytor, 2000.

lates is well known, since Cleveland (1924) showed that the lower termite died in 3 to 4 if the flagellates were removed, but reinfection by them enabled their host to survive. However, more than 70% of termite species belong to the Termitidae, which lack symbiotic flagellates. Lower termites in addition to higher termites are more or less able to produce cellulases by themselves. They secrete end-β-1, 4-glucanases from the salivary gland, while higher termites secrete it from the midgut. The first cellulase gene for the endo-β-1, 4-glucanas was sequenced from a Japanese lower termite, *Reticulitermes speratus*, in 1998.

A biochemical overview of the major processes involved in the symbiosis between termites and their gut microorganisms is shown in Fig. 12. The flagellates ingest wood taken by the lower termites and decompose cellulose as follows.

$$C_6H_{12}O_6 + 2H_2O \rightarrow 2CH_3COOH + 2CO_2 + 4H_2$$

The flagellates provide acetate for termite and bacterial metabolism as well as H_2 and CO_2 for bacterial metabolism.

Celullose in the plant cell wall is combined with lignin to form lignocellulose. Except for fungus-growing termites, the lignin degradation in termite guts remains ambiguous. Some termites such as *Mastotermes darwiniensis* and *Coptotermes acinaciformis* are incapable of digesting lignin, whereas *Nasutitermes exitiosus* seems to be able to degrade lignin at least in part.

D. C/N Balance Symbiosis

Lower and higher termites harbor a high density and diversity of bacteria in the gut: 109 to 1011 cells per ml of gut fluid. Most bacteria are distributed in the paunch of the hindgut, but no clear evidence to support a significant role for bacteria in cellulose hydrolysis has been given. Intestinal bacteria are important in nitrogen fixation, nitrogen enrichment of the diet, pyruvate metabolism, or the metabolism or acetogenic reduction of CO_2.

Carbon elimination through the activities of microorganisms is the most general mechanism of C/N balance. It entails the release of carbon-rich, nitrogen-poor products of digestion leaving an assimilatable fraction of digesta enriched in nitrogen. Gaseous forms of carbon released are principally carbon dioxide and methane, which is formed through respiration, fermentation, and methanogenesis. Almost all termites emit methane and soil-feeding termites tend to emit more methane than wood-feeding termites. Diverse prokaryotes are capable of participating in such transformations.

As long ago as 1925, Cleaveland suggested that termites use nitrogen fixation to explain their survival on a diet low in nitrogen. Nitrogen fixation, which is energy demanding, is only known from bacteria. Nitrogen fixation in termite was first reported by Benemann (1973) and Breznak *et al.* (1973). Tayasu *et al.* (1994) estimated the extent of nitrogen fixation in termites by comparing the natural abundance of ^{15}N in termite tissues with that of their food and air, showing that 30 to 60% of the nitrogen in the body of *Neotermes koshunensis* (Kalotermitidae) was derived from N_2 fixation. However, the variation in the importance of nitrogen fixation in the termite colony is not clear.

Nitrogen is conserved by internal recycling of nitrogenous waste from termites, by recycling of feces, or by cannibalism and nectophagy. Nitrogenous wastes excreted into the gut through the Malpighian tubules are broken down by gut microorganisms and recycled. Termites have symbiotic, uricolytic bacteria, which use uric acid as an energy source, producing ammonia as the major end product. Ammonia provides a readily available source of nitrogen for protein synthesis in many bacteria.

E. Symbiosis with Fungi

Higher plants and animals have developed various symbioses with fungi. The Macrotermitinae of higher termites cultivate the basidiomycete fungus, *Termitomyces*, on fungus gardens (fungus combs) in the nest and are called fungus-growing termites. *Termitomyces* is known only from macrotermitine nests. The fruiting body of *Termitomyces* growing on the comb of *Odontotermes formosanus* at the beginning of rainy season in Okinawa, Japan, is shown in Fig. 13. Ascomycetes of the genus *Xylaria* grow rapidly on the comb, when the comb is removed from the nest.

Fungus-growing habits are known from ants as well (e.g. *Atta*). It is a curious fact, possibly coincidental, that fungus-growing termites, which consume both fungi and fungus combs, are distributed in Africa and southeast Asia, whereas fungus-growing ants, which consume only fungi growing on fungus gardens, are distributed in America.

In *Macrotermes michaelseni* in the grassland of Kenya, old workers (mainly major workers) go out of the nest and collect dead grass pieces. Young workers in the nest consume the pieces and build fungus comb in the mound chambers by depositing fecal pellets of partially digested grass pieces. The comb, which is shaped into round forms resembling brain or sponge, is inoculated with *Termitomyces*, which develop small round white nodules called conidia. Young workers continually deposit the fecal pellets on the top of the comb and consume the conidia, which contain cellulase as well as glucose. Old workers and soldiers consume the old part of the comb from underneath.

The role of *Termitomyces* has been controversial since Konig and Smeathman discovered the fungus in the 18th century. Suggested roles of the fungus are (a)

FIGURE 13 *Termitomyces* (fruit body) growing on the fungus combs of Odontotermes formosanus in Okinawa, Japan. Photo by K. Sugio.

degradation of lignin in the fungus comb, (b) ingesting and utilization of fungal cellulolytic enzymes in conidia to complement its own digestive enzymes, and (c) improvement of nitrogen economy in termite nutrition.

Termitomyces is able to break down lignin in addition to cellulose in the comb to simpler compounds, which are reingested by old termites together with the fungi. In *Macrotermes mulleri*, a synergistic interaction between the cellulolytic enzyme of the termite and fungal origin in the gut is known, while in *M. subhyalinus* and *M. michaelseni*, most (90%) of the glucose requirements of workers can be met by the activity of endogenous intestinal cellulases. The extent to which fungal enzymes contribute to cellulose digestion after ingestion by termites is under debate.

The symbiotic fungi seem to be an important nitrogen source for termites, because nitrogen concentration in the fungi is much higher than that in the fungus comb. Thus, all of the three suggested roles of fungus are partly true. The size and morphology of the combs are different among macrotermitine species, and the role of fungi may vary among them. The comb plays a role as reserved food for termites, especially at times when food is scarce out of the nest.

Alates of *Macrotermes bellicosus* and *Microtermes* spp. carry the spores of *Termitomyces* in their rectal chamber, while *Ancistrotermes cavithorax, A. guinecsis, Macrotermes subhyalinus, Odontotermes pauperus*, and *O. smeathmani* do not carry the fungal spores. They may depend on the source of spores from fungi growing on the soil surface, which the first workers collect.

IV. SOCIAL LIFE

A. Caste Differentiation

A termite colony usually consists of a pair of reproductives and their offspring including workers or pseudergates (false workers), soldiers, nymphs, larvae, and eggs. Termites are hemimetabolous, and their developmental pathway consists of several instars with little morphological change until the final molt to alates. The differentiation into castes such as workers, soldiers, and neotenics occurs when immatures deviate from the pathway leading from egg to winged imago (alate). The more evolved the species, the earlier the caste differentiation.

The most important difference in the caste differentiation exists between the one-piece nest type and separate and intermediate nest types. The one-piece type produces pseuderagtes (false workers), while separate and intermediate types produce workers. The pseudergates, which functionally correspond to workers in separate and intermediate types, retain the capability to be reproductive (alate and neotenic), presoldier, and another pseudergate instar. In this sense, pseudergates are helpers.

The developmental pathways of *Kalotermes flavicollis* of Kalotermitidae (one-piece type) and *Macrotermes michaelseni* of Termitidae (separate type) are shown in Fig. 14. *K. flavicollis* is characterized by flexible or late caste differentiation. On the way to the imago, which develops from the egg through seven molts, two irreversible deviations to sterile soldiers and neotenics and one reversible deviation to pseudergates occur. There is no sexual dimorphism: the sex ratio among pseudergates and soldiers usually depart little from 1.

Caste differentiation is controlled by pheromones. Reproductives (king and queen) secrete sex-specific pheromones that inhibit the metamorphosis of immatures into reproductives, although the inhibitory pheromones have not been chemically identified (Fig. 15). Two inhibitory pheromones are produced by the queen and the king and are passed out of their anuses (proctodeal feeding). When the reproductives die, pseudergates or nymphs molt to neotenics and inherit their parents' colony. The presence of soldiers inhibits the metamorphosis of pseudergates into soldiers, but its physiological basis is not yet known.

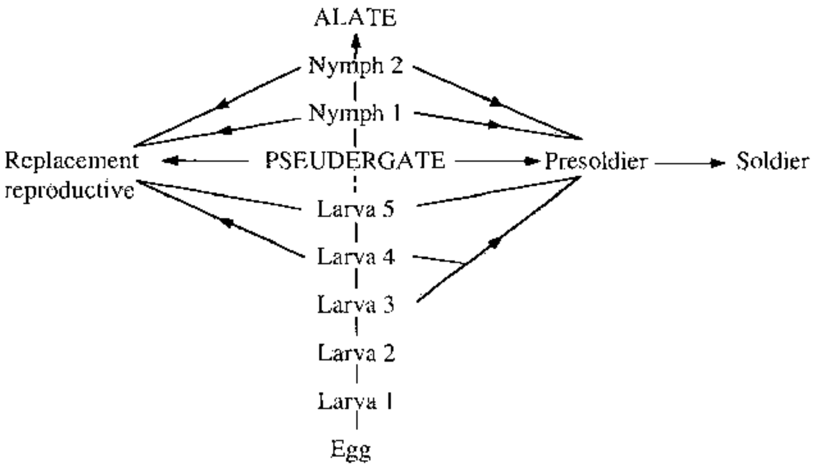
The flexible caste differentiation in *K. flavicolis* is associated with its nesting habits. The Kalotermitidae to which *K. flavicolis* belongs nest in the dead branches of living trees, standing dead trees, and rarely fallen trees, consuming the wood. Those nests are unstable, because wood consumption by nesting termites leads to nest destruction.

In *Macrotermes michaelseni* of the Termitidae, the second instar is an irreversible decision point separating two developmental pathways to imago and sterile workers and soldiers. The number of larval instars is three in the Macrotermitinae and two in the other Termitidae. A sexual dimorphism frequently occurs in workers, while soldiers are of one sex. In *Macrotermes*, major workers are males and minor workers are females, while major and minor soldiers are females. In *Trinervitermes* of the Nasutitermitinae, soldiers are males and workers are females. Replacement reproductives of the Termitidae arise either from alates and nymphs or workers.

B. Evolution of Workers and Soldiers

All species of termites, with some exceptional groups in higher termites, have soldiers. They are adapted to living in the darkness and lack compound eyes. The soldiers of most species are sterile, but fertile soldiers

Kalotermes flavicollis (Lower termite)



Macrotermes michaelseni (Higher termite)

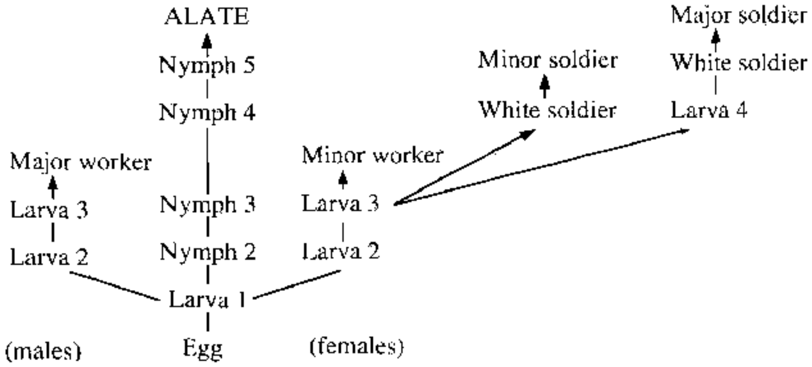


FIGURE 14 Caste development in lower and higher termites. From Roisin, 2000.

that are able to be neotenics are present in *Zootermopsis*, *Archotermopsis* and *Stolotermes* of Termopsidae, a family of primitive lower termites.

Soldiers, which are specialized in the defense of nest and foraging workers, are rich in the variety of morphology. Two types can be distinguished; mandibulate and nasute types. Mandibulate soldiers occur in all families and rely on large, prominent mandibles to defend the nest and colony members by biting or snapping with mandibles. In nasute soldiers occurring only in Nasutitermitinae of the Termitidae, the front of the head is modified to chemical warfare and the mandibles are in many cases reduced. Defense by nasute soldiers are very effective to ants, major predators of termites (Fig. 16). Soldiers are of one size in most species, but are di- or trimorphic in *Schedorhinotermes* of Rhinotermetidae and Termitidae (e.g., *Macrotermes* and *Acanthotermes* of the Macrotermitinae).

The proportion of soldiers to workers (or false workers) in mature colonies varies considerably. Soldiers occupy up to 30% of colony members in the Nasutitermitinae with nasute soldiers, which are usually smaller than workers, while they occupy about 5% in most termites with mandibulate soldiers, which are usually larger than workers. The proportion of soldiers decreases in soil feeders and some of them (mainly Apicotermitinae of Termitidae) lack the soldier caste.

Workers are numerically the largest caste in the colony and consist of sterile males or females which are wingless, in most cases unpigmented, and lack special external modifications. Compound eyes are usually absent. Workers provide labor for gathering food, feeding dependant castes (larvae, soldiers, nymphs, and reproductives), tending eggs, and repairing and enlarging the nest and gallery system. Worker polyethism is based on age and sex; old workers tend to predominate in

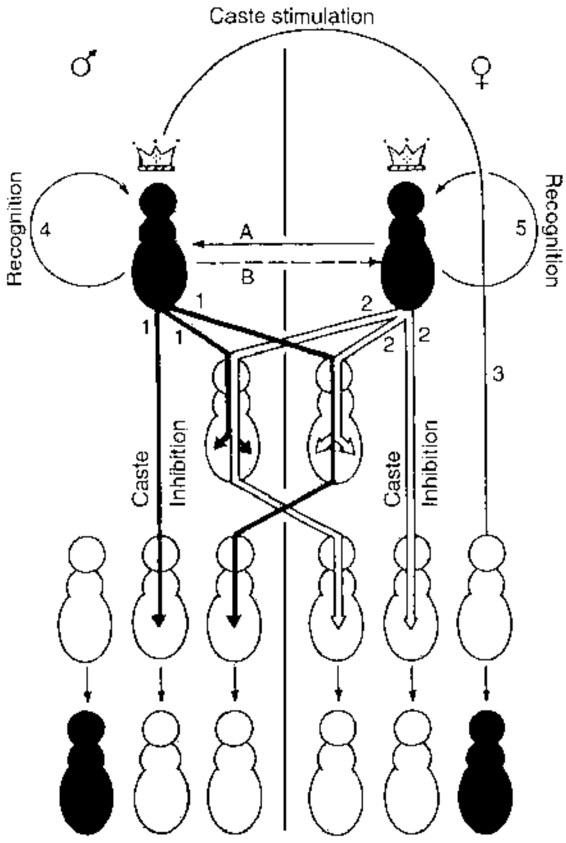


FIGURE 15 The pathways of pheromone action in the control of reproductive caste formation in the termite Kalotermes flavicollis. In the top row the pair of "crowned" figures represents the reproductives: the functional male (king) to the left and the functional female (queen) to the right. The remaining figures all represent pseudergates. The king and queen produce substances (labeled 1 and 2, respectively) that inhibit transformation of pseudergates into their own royal castes. These inhibitory pheromones are passed directly from the reproductives to the pseudergates and are also circulated indirectly through the digestive tracts of the pseudergates. Another male substance (pheromone 3) stimulates the female pseudergates to change into the reproductive caste, but the reverse relation does not hold. When supernumerary royal males are present, they recognize each other (through pheromone 4) and fight; similarly, supernumerary royal females recognize each other (through pheromone 5) and fight. Finally, royal males stimulate production of pheromone 2 in royal females, and royal females stimulate production of pheromone 1 in royal males; the nature of the stimuli labeled A and B is unknown. From Wilson, 1971.

outer tasks, while young workers tend to predominate in inner tasks. Workers are of one size in most species, but are dimorphic in some species. Major workers tend to engage in outer tasks, while minor workers tend to engage in inner tasks.

The presence or absence of sterile castes is shown in relation to life type on the phylogenetic tree in Fig. 17. Sterile soldiers appeared earlier than sterile workers, evolved once in one-piece type prototermites with false workers, and disappeared secondarily in some soil-feeding termites. On the other hand, the sterile worker caste has evolved polyphyletically. Separate and intermediate type termites (Mastotermitidae, Hodotermitidae, most Rhinotermitidae, and Termitidae) possess a worker caste, whereas one-piece type termites (Termopsidae, Kalotermitidae, and *Prorhinotermes* of Rhinotermitidae) possess a false worker caste. The sterile worker may have evolved from the false worker at least three times independently in accordance with the change from one-piece nest to intermediate and separate types.

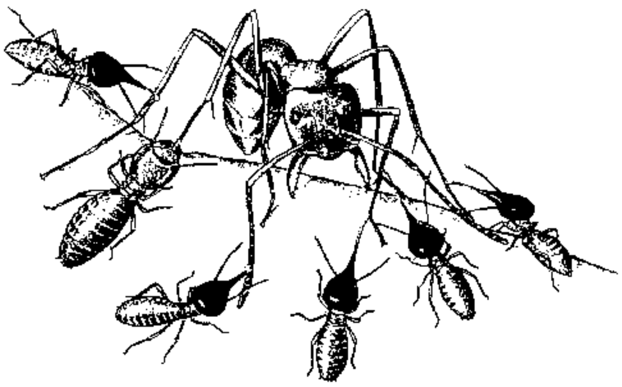


FIGURE 16 Defensive posture of a worker and nasute soldiers of *Nasutitermes* against an ant worker. From Behnke, 1977.

C. Diversity in Social Life

The social organization of termites is similar to that of social Hymenoptera in that the colony is organized based on the division of labor due to caste and age. However, there is an important difference between them. In contrast to Hymenoptera, whose active members are imago (adult, which are suppressed to be reproductive), active members such as workers and soldiers of termites are individuals which are suppressed to be imago.

The degree of eusociality, fecundity of reproductives, and the other life history characteristics are diverse among termite species, and they are summarized in relation to life type in Table IV. The reproductive skew, defined as the probability that workers and soldiers forego direct reproduction, is a good indicator of eusociality (0: all colony members reproduce; 1: colony members forego direct reproduction except a single colony breeder). The skew is low in the one-piece type, where pseudergates have the potential to be reproductive, while it is high in the intermediate or separate type, where workers have little chance to become reproductives.

No. of species	411	1	20	15	305	1879
Distribution	WW	Au	Amp	Af,As	PT* WW	PT
Nest site	W	W,S	W	S	W	W,S,E,A
Food	W	W	W	DG	W	W,DG,DL,S,Li
Symbionts	P,B	P,B	P,B	P,B	P,B	B,F
Colony size	10^3	10^6	10^2	?	10^4* 10^6	10^6
Life type	O	I	O	S	O* I,S	S
Sterile caste	S	S,W	S	S,W	S* S,W	S(-),W

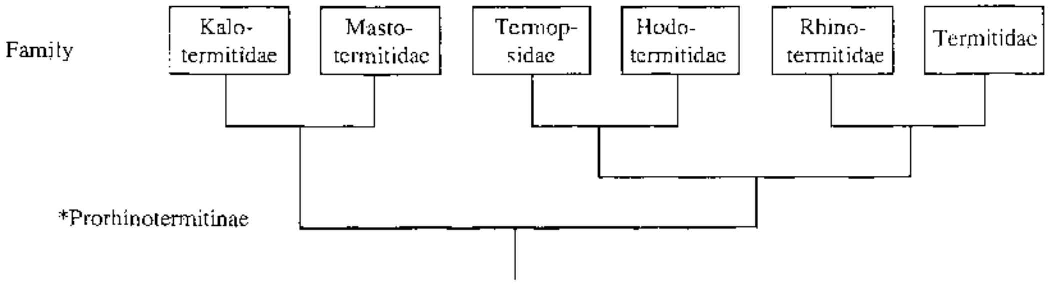


FIGURE 17 Six families of termites and a summary of their characterization in terms of distribution, nest site, food, symbionts, colony size, life type, and sterile caste. Distribution: Au, Australia; WW, worldwide; Af, Africa; As, Asia; Amp, amphitropical; PT, pantropical. Nest site: W, wood; S, subterranean; E, epigeal mound; A, arboreal. Food: W, wood; DG, dead grasses; DL, dead leaves; S, soil; Li, lichen. Symbionts: P, Protozoa; B, bacteria; F, fungi. Life type: O, one-piece; I, intermediate; S, separate. Sterile caste: S, soldier; W, worker. From Higashi and Abe, 1997.

The reproductive skew is negatively correlated with the extent of noetenic production. The percentage of genera producing neotenics is high in the one-piece type (100% in Termopsidae and *Prorhinotermes* of the Rhinotermitidae) except the Kalotermitidae (48%), medium (100% in Mastotermitidae and 71% in Rhinotermitidae) in the intermediate type, and low (67% in Hodotermitidae and 0–20% in Termitidae) in separate types.

A remarkable difference among life types is the fecundity of primary reproductives. It is low in the one-piece type, whereas it is sometimes extremely high in the separate type (e.g., a queen of *Macrotermes subhyalinus* lays 36,000 eggs per day).

In comparison with most insects, primary reproductives survive for a long time—10 to 12 years for *Incisitermes minor* (one-piece type) and 20 to 50 years in Macrotermitinae (separate type). It is noteworthy that the longevity of reproductives is similar to that of pseudergates in the one-piece type, whereas it is much longer than that of workers in the separate type.

The age at which termites begin to reproduce is early (beyond the second year in *Kalotermes flavicollis*, beyond the fourth year in *Incisitermes minor*, after the fourth year in *Zootermopsis*, after the sixth year in *Neotermes tectonae*) in the one-piece type, and late (after 5 to 10 years in many species of Rhinotermitidae and Termitidae) in intermediate and separate types. The size of mature colonies producing alates ranges from a few hundred individuals for Kalotermitidae (one-piece type), many thousands or much more in Mastotermitidae, Rhinotermitidae (intermediate type), and Termitidae (separate type).

The colony size is shown in Table V, although the accuracy of estimation is different. The one-piece type such as the Kalotermitidae and Termopsidae is small, usually up to 10,000, while the separate type such as mound builders of the Termitidae, in particular *Macrotermes* and *Nasutitermes*, is very large up to some millions. The intermediate type, including the Rhinotermitidae and Mastotermitidae, is sometimes very large up to some millions.

TABLE IV

Distinguishing Social Characteristics Associated with Each Nest Type

Life type.	One piece type	Intermediate type	Separate type
Termite group	Termopsidae, most Kalotermitidae and *Prorhinotermes* of Rhinotermitidae	Mastotermitidae, few Kalotermitidae, most Rhinotermitidae and some Termitidae	Hodotermitidae, some Rhinotermidae and most Termitidae
Reproductive skew'	0–0.5	0.4–0.75	0.7–1.0
Task specializations	Sterile soldiers, false workers and age polyethism	Sterile soldiers, sterile workers and age polyethism	Sterile soldiers, sterile workers and age polyethism sterile soldiers (except in some soil feeders), sterile workers and age polyethism
Fecundity of primary reproductive (eggs/day)	Small *Cryptotermes havilandi*: 8		Large *Odontotermes obesus*: 26,000–86,000 *Macrotermes subhyalinus*: 36,000 *Nasutitermes surianamensis*: 3,900 *Cubitermes severus*: 50–600
Age of primary reproductives	Field-observed reproductive 4–5 years in *Zootermopsis* spp. and 10–12 years in *Incisitermes minor*	*Mastotermes darwiniensis* lab-reared reproductives up to 17 years; *C. formosanus* at least 9 years; *R. lucifugus* primaries and neotenics up to 7 years in the lab	Primary reproductives 20 years; queens 20–25 years in Macrotermitinae spp. (estimates based on size of physogastric queen)
Age of workers and soldiers	*Zootermopsis* false workers up to 4 years, soldiers 4–5 years in the lab; *Neotermes isularis* soldiers and false workers more than 6 years in the lab	*Reticulitermes lucifugus* soldiers up to 5 years in the lab; *R. hesperus* workers 3–5 years in the field; *Coptotermes acinaciformis* workers and soldiers 4 years in the lab	*Macrotermes* spp. workers and soldiers less than a year in the field and lab; *Macrotermes natalensis* up to 1.5 year in the lab; *Cubitermes ugandensis* worker 196–339 days
Colony foundation	A male and female dealate pair	A male and female dealate pair by a cohort of neotenics and workers	A male and female pair (sometimes multiple reproductives) or sociotomy or secondary dealates reproductives, brachypterous neotenics with workers, rarely ergatoid reproductives
Dispersal distance of alates	Short *Cryptotermes*: 1–45m *Kalotermes*: 20–50m *Incisitermes*: 120m *Zootermopsis*: 300m	Short *Reticulitermes*: 10–200m	Long *Odontotermes*: over 0.8km *Macrotermes*: a few km

After Shellman-Reeve, 1997.

D. Organization of Foraging

Foraging of termites has evolved in concert with the change of life type. The one-piece type of termites do not forage out of nest, and the nest site selection by alate reproductives plays a role of food research. Dealate reproductives of *Zootermopsis* (Termopsidae) look for a colony founding site in the wood cambium layer containing nitrogen of relatively high concentration.

In intermediate and separate type nesters, a group of workers and soldiers go out of the nest and search for new food sources. Many species show aggressive behavior toward the other colony members of the same or different species, and the foraging territory, which is intra- and interspecifically defended, is known from many termite species.

The division of labor in foraging activities is based on caste and age. Soldiers usually defend foraging workers, and in some species they explore new areas for food and recruit workers to new food sites. Workers mainly collect and process food, and they also defend foraging territory.

Foraging activities are mediated by trail pheromones emitted from the sternal gland. Several sub-

TABLE V

Characters Showing Evolutionary Change within Termites

Characters showing advance	
Primitive condition	Derivative condition
Behavioral-physiological	
Small numbers in the nature colony	Large numbers in the mature colony
Excavated nests with little construction	Elaborately constructed carton nests
Relatively little care of eggs, nymphs, and adults	Relatively great care of eggs, nymphs, and adults
Food not store	Food stored
Nutritive dependence upon symbiotic intestinal flagellates	Nutritive independence from symbiotic intestinal flagellates
Damp-wood eaters	Dry-wood eaters
Wood eaters	Deaf leaf and grass eaters
Wood eaters	Humus eaters
No fungus gardens	Fungus gardens
Few or no termitophiles	Termitophiles, often species-specific and highly modified
Workers forage only within excavated tunnels in wood	Workers forage outside nest, sometimes in covered tunnels and sometimes on exposed odor trails
Reproductive castes	
Small abdomen of queen with small ovaries and glandular tissue	Large abdomen of queen with large ovaries and glandular tissue
Capacity to produce substitute (neotenic) kings and queens	Reproduction confined to primary reproductive caste (image)
Frontal gland absent	Frontal gland present
Front coxae smooth	Front coxae with ridge or projection
Workers	
False worker caste present	True worker caste present
Soldiers	
Head elongate, large, smooth, somewhat flattened, with large, curved, and toothed biting mandibles	Head round, or extremely flat, or phragmotic; surface rough or with ridges and projections, mandibles elongate and thin, or twised for snapping, or with proejction for frontal gland
Mandibles with smooth cutting edges or with reduced teeth	Mandibles with serrated cutting edges
Pronotum wide, flatly convex, and with smooth edges	Pronotum narrow, saddle-shaped and sometimes with serrated edges
Front coxa smooth	Front coxa ridge or projection

Characters showing regression	
Primitive condition	Derivative condition
Imagoes (primary reproductives)	
Y-suture of head present	Y-suture of head reduced or absent
Two well-developed ocelli present	Ocelli reduced or absent
Antennae with numerous segments	Antennae with fewer segments
Mandibles with 2 or 3 prominent marginal teeth with sharp basal notches	Mandibles with reduced teeth and notches
Five tarsal joints present	Second tarsal joint reduced or absent
Arolium present between tarsal claws	Arolium absent
Carci with numerous segments (up to 8)	Cerci reduced, usually with 2 segments
Styli present	Styli absent
Egg mass in cluster	Eggs laid separately
Genitalia similar and clearly homologous to those of cockroaches	Genitalia reduced or absent
Hind wing with anal lobe	Hind wing without anal lobe
Pronotum wide and flat	Pronotum narrow and saddle-shaped
Workers	
Compound eyes faceted and pigmented	Compound eyes reduced or absent
Mandibles as in primitive imagoes	Mandibles as in derivative imagoes
Soldiers	
Antennae with many segments (up to 29)	Antennae with fewer segments (As few as 10)
Compound eyes pigmented and faced	Compound eyes unpigmented, or nonfaceted or even absent
Ocellus spot present	Ocellus spot absent
Mandibles with 2 or 3 large marginal teeth	Mandibles with reduced marginal teeth
Soldiers present	Soldiers absent

These changes have occurred at various places in termite evolution. Modified from Wilson, 1971.

stances are isolated from the gland: n-hexanoic acid in Termopsidae, dodecatrienol in Rhinotermitidae, and neocembrene in Nasutitermitinae of the higher termites. Many chemicals that are not endogenous to termites cause an orientation response. 2- phenoxyethanol from the ink of certain ball-point pens have a trail-following effect on *Coptotermes formosanus*.

Territoriality and elaborate division of labor in foraging were studied well in the Macrotermitinae and the Nasutitermitinae of higher termites, because some species of them forage on the ground. A fungus grower, *Macrotermes michaelseni* in Kenyan savanna, constructs a large mound and has a very large foraging territory based on subterranean galleries that extend up to 50 m from the mound, run horizontally, and lead to foraging holes on the ground surface. The species possesses four sterile castes, major and minor soldiers, and major and minor workers. Major soldiers mainly defend the mound, while minor soldiers mainly defend foraging workers. Major workers are foraging specialists, while minor workers mainly perform intranidal works. The division of labor between major and minor workers is more or less different among species of *Macrotermes*.

During the night, major workers emerge from the foraging hole, cut off dead grasses, fallen twigs and cow dung on the ground, and bring them into the foraging hole. They form foraging columns (up to 1 m) between the foraging hole and the foraging site, both sides of which are guarded by minor soldiers and rarely major soldiers. The column is attacked by various species of ants. The foraging areas of *Macrotermes carbonarius*, a mound builder in southeast Asia, change irregularly but seem to be arranged to avoid redundant foraging in the long term (Fig. 18).

Black marching termites of *Hospitalitermes* of the Nasutitermitinae in southeast Asia form soldier-led foraging columns reaching up to 100 m in length from the nest at the base of trees to feeding sites on standing tree trunks where workers collect lichens. Soldiers explore the feeding site and recruit workers to the feeding site, defending the workers. Three worker castes show division of labor during food collection; minor workers gnaw at lichens and form a food ball that is passed on medium and major workers, which carry the food ball to the nest. Medium workers also form food balls.

Two arboreal species of Nasutitermitinae, *Nasutitermes nigriceps*, and *N. nigriceps* form a clear foraging territory in a Panamanian mangrove forest. They build arboreal nests and forage along carton-convered tunnels on branches, which radiate from the nest, and have a nonoverlapping foraging area that is intra- and interspecifically defended.

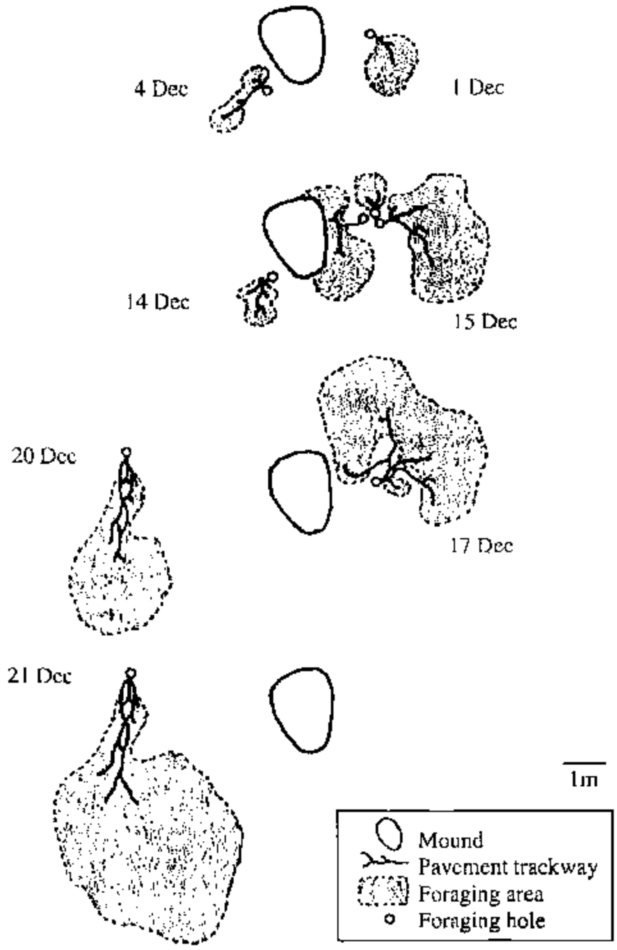


FIGURE 18 Daily change of foraging areas and pavement trackways of a fungus-growing termites, M. carbonarius in Thailand. From Sugio, 1995.

E. Life History of Fungus-Growing Termites

A new colony is founded by dealates or budding of the parent colony. In most cases, a pair of dealates found a new colony to be primary reproductives. Multiple alate-derived queens are restricted to the Termitidae (ca. 40 species are known). Colony budding is achieved by active migration of swarms containing all castes or by division of a diffusely organized colony. The former, called a "sociotomy," is known only in the Termitidae (e.g., *Trinervitermes bettonianus*, *T. gratiosus*, *Syntermes territus*, and *Anoplotermes* spp.). The latter is known in Mastotermitidae, Kalotermitidae, Termopsidae, Rhinotermitidae, and Termitidae, where the extension of nests is coupled of production of supplementary reproductives.

The life cycle of *Macrotermes michaelseni*, a socially

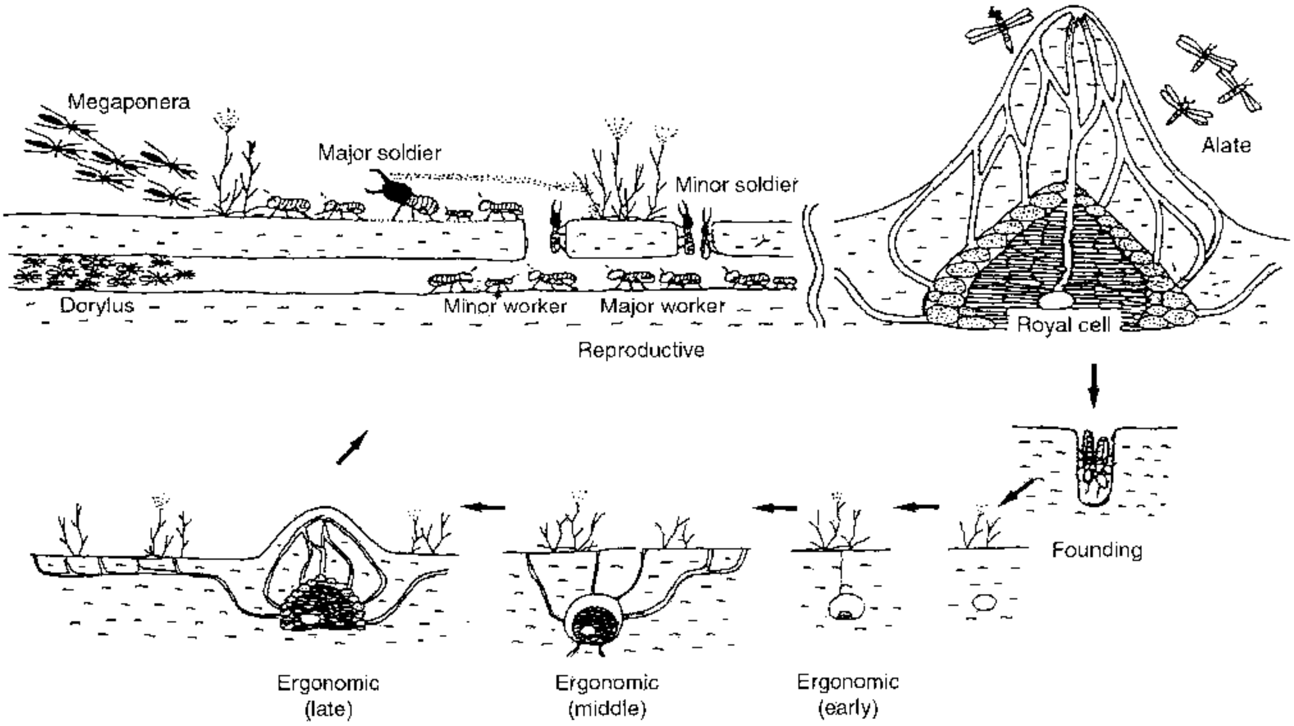


FIGURE 19 Life cycle of the fungus-growing termite, *Macrotermes michaelseni*, in the grassland of Kenya. From Abe, 1995.

advanced fungus-growing species of the Termitidae in the grassland of Kenya, is shown in Fig. 19. After the nuptial flight at the beginning of rainy season, a male (king) and female (queen) pair of dealates form a tandem, dig a small cell in the soil, copulate, and initiate a subterranean incipient colony. Occasionally three or more dealates found a new colony together. The primary queen begins to lay eggs within a few days after mating. They hatch after some weeks.

The first brood is tended by the king and queen and is differentiated into soldiers as well as workers. The first workers begin foraging at about 10th week after colony founding in the fungus grower *Microtermes*. The king and queen are similar in body size at colony founding, but the abdomen of the queen becomes much larger than that of the king with time.

Within a few years, *M. michaelseni* constructs a mound that reaches 3 m in height at the mature stage. A large mound builder has a territory based on subterranean galleries, which run horizontally in a shallow layer (10–15 cm). Young subterranean colonies can survive within the territories of large mound builders, because they are located at a deeper layer (20–40 cm) than the subterranean galleries of large mound builders and are able to avoid their attack. However, with colony growth they construct foraging galleries in the shallow layer and are killed by large mound builders. Thus in *Macrotermes michaelseni*, large mounds tend to be uniformly distributed due to territoriality, and small (young) mounds are rare in Kenya.

On the other hand, both large and small mounds of *M. bellicosus* are distributed randomly in Nigerian savanna. The mortality rate of large mound builders of *M. bellicosus* is very high due to predation by doryline ants and aardvarks. Thus, young subterranean colonies are not seriously affected by the proximity of mound builders and survive to construct small mounds, which are randomly distributed.

A marked increase in the number of small mounds following the death of large mound-building colonies due to the severe drought has been reported *M. bellicosus* in Senegal and *M. subhyalinus* in Kenya. Thus, when the mortality rate of large mounds is low, they show uniform distribution due to their territoriality, whereas when their mortality rate is high, young mounds are continuously added and mounds as a whole tend to show random distribution. The duration of mound occupation has been poorly understood: 80 years for *Macrotermes* in Africa, more than 100 years for a large mound of *Nasutitermes triodiae* in Australia, and 20 to 40 years for *Amitermes vitiosus* in northern Australia.

V. THE GLOBAL DIVERSIFICATION OF TERMITES

A. Major Events in Termite History

Three major events of global diversification that Isoptera has experienced are identified (Higashi and Abe, 1997): (a) the evolutionary radiation of original termites, (b) the expansion and diversification of separate type termites and the diversification of one-piece type termites driven by the expanding separate type termites into fragmented habitats, and (c) the radiation of higher termites.

As stated earlier, Cretaceous fossil species belonging to Hodotermitidae, Termopsidae, and Mastotermitidae are known from Europe, Asia, and North and South America. This shows that primitive lower termites had been globally distributed and diversified in the Cretaceous. Therefore, the first event, the evolutionary radiation of original termites, must have occurred at latest in the Jurassic and Cretaceous periods of the Mesozoic. All living species of Hodotermitidae are specialized in grass feeding, while living species of Termopsidae and Mastotermitidae are wood feeders. Grass feeding habits of the Hodotermitidae must have evolved in the Tertiary period, and Cretaceous Hodotermitidae must have consumed wood, because grasses evolved in the Tertiary.

The fact that the oldest fossil record is the worker of Hodotermitidae is interesting, because this suggests that the evolution of the worker caste occurred in the Mesozoic. We do not know if Cretaceous termites of Hodotermitidae nested in wood or soil.

Many fossil species of the Kalotermitidae whose living species are mostly of the one-piece type have been found in the Tertiary period. It is probable that the second major event, the expansion and diversification of separate type termites and the diversification of the one-piece type termites driven by the expanding separate type termites into fragmented habitats, occurred in the Tertiary.

Only a few fossil species of the higher termites are known from the Tertiary. The third event, the radiation of higher termites, may have occurred in the late Tertiary or the Quarternary. On the other hand, the worldwide distribution of higher termites suggests their early evolution and dispersal before the breakdown of Gondwanaland in the Cretaceous. As shown later, it is quite difficult for termites to cross the oceans. It is probable that the radiation of higher termites may have occurred in the late Tertiary or the Quarternary, although the origin of higher termites may be before the breakdown of Gondwanaland in the Cretaceous.

B. Dispersal Ability of Termites

Islands sterilized by volcanic eruption are good for studying the process of colonization. In 1883, Krakatau, located about 60 km from Java and 40 km from Sumatra, erupted violently and more than half of the island disappeared, leaving the remaining Krakatau and two neighboring islands. The entire flora and fauna of the islands were almost destroyed, but they were rapidly colonized by organisms of Java and Sumatra.

In the Krakatau Islands, three, three, and seven species of termite were collected in 1908, 1919–1933, and 1982, respectively. They were characterized by the dominance of the lower termites (Kalotermitidae and Rhinotermitidae), the one-piece type, and the absence of soil feeders, although a wood-feeding arboreal species of *Nasutitermes* of the higher termites was found in the Krakatau Islands. On the other hand, the supposed source area (the tip of West Java) was characterized by higher species diversity (13 species), the dominance of higher termites and separate type, and the presence of soil feeders. In the tip of West Java, the one-piece type termites of the Kalotermitidae are confined to the coastal forest, while soil feeders are confined to inner forests.

Alates of termites are specialized for dispersal within a zone of calm air near the ground. The dispersal distances are short in lower termites (up to 300 m) and a little long in higher termites (a few kilometers). It is not impossible but difficult for termites to disperse widely or to reach remote islands by flight. Therefore, termites usually cross the sea by rafting. The Kalotermitidae, the first colonizers, have some preadaptive attributes for oversea dispersal: (a) they tend to be confined to the coastal area and may be easily swept into the sea, (b) they are more tolerant to sea water than other types, and (c) their caste differentiation is flexible.

On the other hand, it is difficult for soil-feeding termites to cross the sea because they have little opportunity to be in rafting wood. Thus the dominance of the Kalotermitidae (one-piece type) and the Rhinotermitidae (intermediate type) in many islands is explained in terms of the differential dispersal ability of termites.

C. The Evolutionary Radiation of Original Termites

The first event, the evolutionary radiation of the original termites into a new habitat of wood throughout the world, may have been driven by an efficient utilization of wood as an abundant and stably supplied food resource and nest substrate, which became possible by the evolution of two symbioses for termites (i.e., the cellulose digestion symbiosis with cellulolic flagellates and the C-N balance symbiosis with nitrogen fixing bacteria). All descendants of the Mesozoic families of Hodotermitidae, Termopsidae, and Mastotermitidae harbor symbiotic flagellates and nitrogen-fixing bacteria in the gut.

Once a termite obtains a means for digestion and C-N balance through association with microorganisms, then wood, which is superabundant but extremely hard to digest and C-N unbalanced, becomes a "well-protected" food resource that can be monopolized by the termite. The diversification of termites into the Hodotermitidae, Termopsidae, and Mastotermitidae together with their worldwide distribution in the Cretaceous may be viewed as an example of niche opening, which often leads to an adaptive radiation.

Conquering the wood habitat should promote the evolution of a false worker caste, because life in a piece of wood causes (a) greater need for concentrating nitrogen, (b) higher possibility of nest succession due to the longer duration time of the nest relative to the longevity of reproductives, and (c) lower possibility of finding an alternative nesting place due to the heterogeneity of resource (wood) distribution. The evolution of this subsociality should further have enhanced the opportunity for eusociality—that is, the production of sterile soldier caste.

D. Separate Type Expansion and One-Piece Type Diversification

The second major event came with the development of separate type termites, which separate their feeding sites from their nest. They may have expanded, replacing the existing one-piece type termites through a competitive exclusion process, the detail of which was examined by Higashi and Abe (1997). This is a consequence of asymmetrical competition between one-piece and separate types due to the fact that a piece of wood is food and nest for the one-piece type but only food for separate type termites. On the other hand, some groups of one-piece type termites, which were driven out by the expanding separate type termites into fragmented habitats, should have gone through a diversification process.

This process could be reconstructed based on the present peculiar distribution patterns of one-piece type termites and separate type termites. The three groups of one-piece termites show quite different marginal distribution patterns. Damp wood termites (all species of Termopsidae) show a clear amphitropical distribution (Fig. 20). Ordinary wood termites of *Prorhinotermes* are widely distributed in tropical and subtropical regions but tend to be confined to islands. Dry wood termites (Kalotermitidae) are also widely distributed in the tropical and subtropical regions but tend to be confined to dry dead parts of standing trees.

On the other hand, separate type termites, represented by the Termitidae, are widely distributed in tropical mainlands, the central areas of termite distribution. Intermediate type termites, represented by Rhinotermitidae, are widely distributed from tropical to temperate regions. Thus the distribution of the one-piece type and separate type is complimentary, while that of intermediate type covers the ranges of both types.

Although a full explanation of such distribution patterns requires one to take into consideration other factors such as plate tectonics and global change of climate, ecological mechanisms, specifically interspecific interactions and differential dispersal ability, have been a major force in determining the ranges of distribution.

The asymmetrical competition theory predicts that as the latitude (ocean) gets higher (closer), the resource supply rate of the one-piece type increases relative to that of the separate type, because more wood is available to the one-piece type due to the slower decay rate of damp (dry) wood at higher latitude (seacoast) areas under a given predation pressure. Thus, the one-piece type may compensate its disadvantage from its asymmetrical competition with the separate type and becomes dominant at higher latitude (seacoast) areas. This explains the questioned marginal distribution patterns of damp and dry wood termites. The distribution of ordinary wood termites confined to islands may simply be explained as a result of escape from asymmetric competition due to their dispersal ability.

Separation of feeding site from nest (reproductive site) should enhance the C-N balance symbiosis by adding the option of selective carbon elimination and promoting the evolution of a true worker caste. These should further have enhanced the superiority of the separate type in resource utilization, thus driving the expansion and diversification of separate type termites.

Among the one-piece type termites driven out by the expanding separate type termites, those groups driven into fragmented habitats along seacoasts and islands are expected to have gone through a diversification process. In fact, dry wood termites show a high species diversity as a family.

E. The Radiation of Higher Termites

The third major event, the radiation of higher termites that are free from flagellates, may have been driven by the advantage that a termite would have if it could digest cellulose efficiently enough and eliminate cellu-

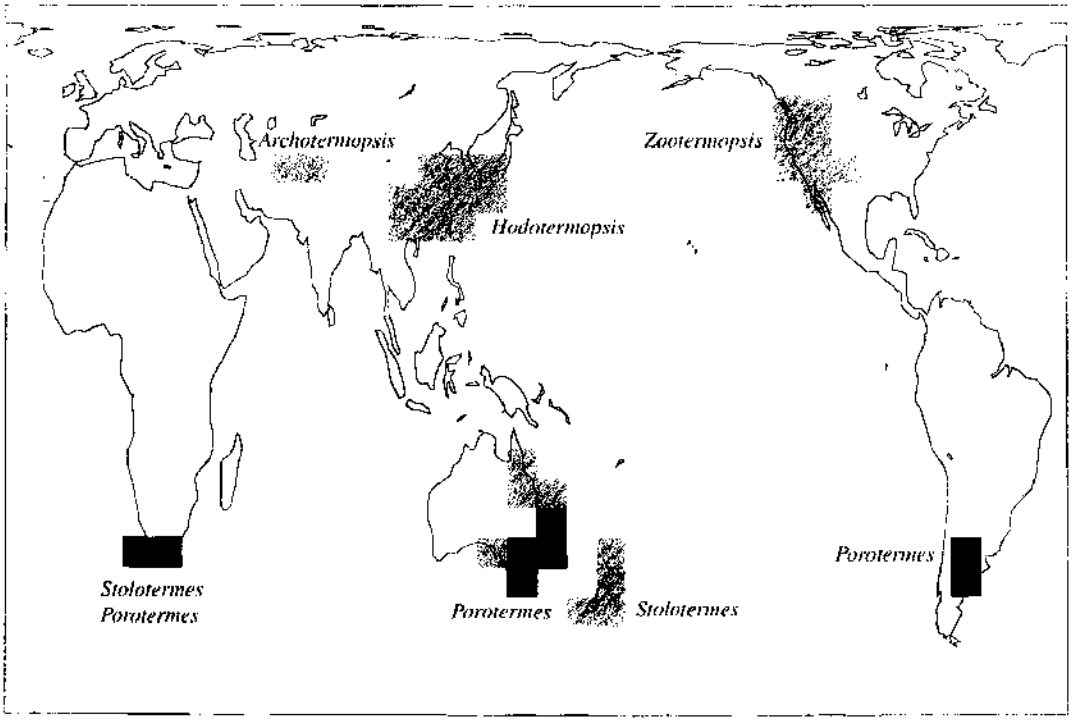


FIGURE 20 Distribution of the Termopsidae. From Eggleton, 2000.

lolic flagellates, which became possible by eliminating the lignin cover, the major obstacle in cellulose utilization, perhaps through the shift of their diet from wood to a food with less lignin cover.

Those termites that became higher termites must have obtained their own cellulase activity. The present-day termites have their own cellulase activity. What prevents the elimination of cellulolic flagellates from lower termites may be the inefficiency of the termites in cellulose digestion, which should cause the termites a shortage of (not glucose but) energy supply. Thus, cellulolic flagellates are still symbionts to lower termites, though they are not for cellulose digestion but for energy supply. Higher termites are those that succeeded in increasing the efficiency of cellulose digestion to the extent that they no longer need help from cellulolic flagellates for energy supply either. This may be the driving force for their radiation.

For increasing the efficiency of cellulose digestion, the elimination of the lignin cover is expected to be the most effective. An obvious means to achieve the elimination of the lignin cover is to shift the diet from wood to a food with less lignin cover. The nest-food separation makes possible the selection of food. Thus, it is expected that higher termites may appear only from separate type termites, which is indeed the case.

Three major groups of higher termites are recognized: nasute termites (Nasutitermitinae), soil (humus)-feeding termites, and fungus growing termites, although there is some overlap among them. They seem to have succeeded in the elimination of the lignin cover.

The Nasutitermitinae show the broadest distribution and the highest species diversity among the higher termites. Their radiation was driven by the evolution of a new type of soldier with a chemical weapon against ants, their major predator, which made it possible to forage out more freely and select food with less lignin cover. The radiation of soil feeders was driven by the shift of diet from wood to humus, which contains less lignin. The radiation of fungus-growing termites was driven by broadening the diet and a new symbiosis with fungi for C-N balance and for lignin reduction.

VI. TERMITES IN ECOSYSTEMS

A. Terrestrial Ecosystems

Terrestrial ecosystems have some characteristics different from aquatic ecosystems—for example, the dominance of the detritus chain among grazing, detritus, and microbial chains; the rare cascading effect of predators on vegetation; the abundance of fungi; and so on. This is partly because vascular plants in terrestrial ecosystems produce lignin, an organic matter extremely resistant to chemical degradation. Plants cover cellulose with lignin to form lignocellulose as a major component of the cell wall. Cellulose is an energy source for many heterotrophs, but its availability decreases markedly when covered with lignin.

Lignin makes it possible for terrestrial plants to support leaves horizontally and transport water from root to leaves and branches, providing various heterotrophs with heterogeneous habitats. Furthermore, lignin regulates the decomposition process of dead plant material, and its remnants play an important role on the formation of humus associated with soil fertility.

Therefore, lignin is a key substance for understanding the architecture and the material flow in terrestrial ecosystems. Considering that microorganisms, especially fungi, are major decomposers of lignocellulose, fungi are reasonably abundant in terrestrial ecosystems, whereas they are poor in aquatic ecosystems that lack lignocellulose.

Brown rot among fungi mainly decomposes cellulose, whereas white rot decomposes both lignin and cellulose. Most soil animals are not able to decompose lignin chemically but decompose it physically and promote the chance of contact between lignocellulose and microorganisms. Termites are distinct among soil animals in that some of them decompose lignocellulose efficiently with the aid of microorganisms in the gut or in the nests.

Many termites live in lowland tropical forests, where 50 to 80 species are found in 1 ha area in Fig. 21. Their food is rich in variety including dead branches of living trees, fallen logs, fallen leaves, lichens, and soil. Those termites are functionally largely divided into two groups: soil feeders and litter feeders. Litter feeders are furthermore divided into lignocellulose feeders and cellulose feeders.

Soil feeders, confined to three subfamilies of higher termites, are concerned with the formation of humus

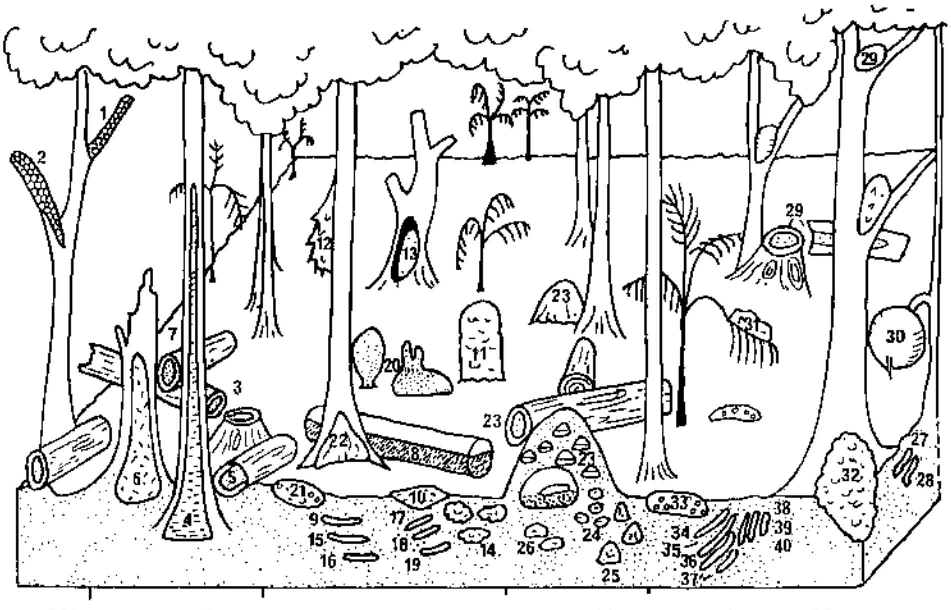


FIGURE 21 Termites in lowland dipterocarp forest in West Malaysia. Kalotermitidae. 1. *Glyptotermes*; 2. *Neotermes*; Rhinotermitidae: 3. *Termitogeton*; 4. *Coptotermes*; 5. *Parrhinotermes*; 6. *Schedorhinotermes*; 7. *Heterotermes*; Termitinae: 8. *Protohamitermes*; 9. *Prohamitermes*; 10. *Labritermes*; 11. *Globitermes*; 12. *Microcerotermes*; 13. *Amitermes*; 14. *Procapritermes*; 15. *Pericapritermes*; 16. *Coxacapritermes*; 17. *Oriencapritermes*; 18. *Syncapritermes*; 19. *Mirocapritermes*; 20. *Dicuspiditermes*; 21. *Homallotermes*; 22. *Termes* (in *Hospitalitermes* nest); Macrotermitinae: 23. *Macrotermes*; 24. *Microtermes*; 25. *Odontotermes*; 26. *Hypotermes*; Nasutitermitinae: 27. *Hirtitermes*, 28. *Havilanditermes*; 29. *Nasutitermes*; 30. *Bulbitermes*; 31. *Lacessititermes*; 32. *Hospitalitermes*; 33. *Longipeditermes*; 34. *Leucopitermes*; 35. *Aciculitermes*; 36. *Aciculioiditermes*; 37. *Subulioditermes*; 38. *Malaysiotermes*; 39. *Proaciculitermes*; and 40. *Oriensubulitermes*. From Mabberley, 1992.

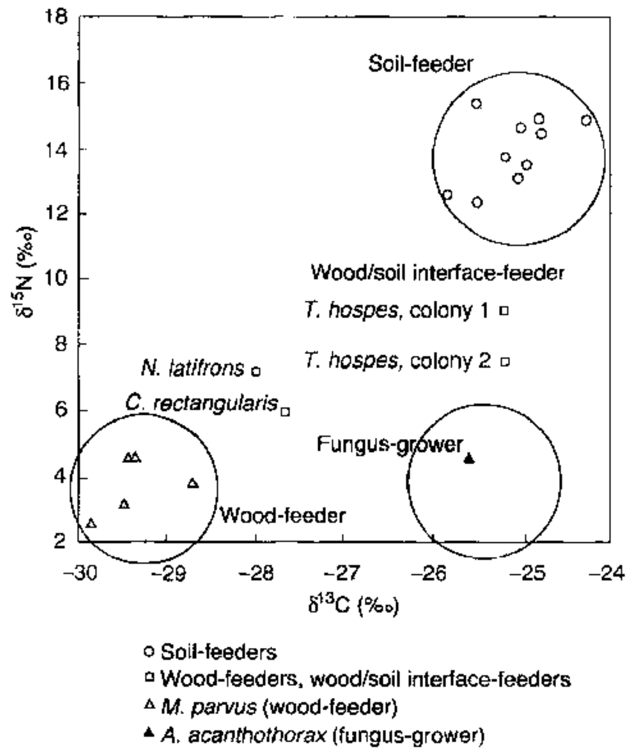


FIGURE 22 Plot of mean $\delta^{15}N$ against mean $\delta^{13}C$ for the tissues of worker caste termites from the Mbalmayo Forest Reserve, Cameroon. Soil feeders (colonies of eight species, open circles) cluster at the top right and colonies of the wood-feeding *Microcerotermes parvus* (open triangles) at lower left. Species feeding on highly decayed wood (*Cephalotermes rectangularis*; *Termes hospes*) are intermediate. A single fungus grower, *Acanthotermes acanthothorax*, and the wood-feeding *Nasutitermes latifrons* are also shown. From Tayasu *et al.*, 1997.

associated with soil fertility. Lignocellulose feeders (fungus growers of Macrotermitinae) decompose both lignin and cellulose rather completely with the aid of the white rot, *Termitimyces*. As they do not leave many lignin remnants, they may functionally well be called giant moving white rot. Cellulose feeders, including most of the other termites, decompose cellulose in dead plant material, leaving lignin remnants. They may functionally well be called giant moving brown rot. Roughly saying, soil feeders, wood feeders, and lignocellulose feeders occupy about 45%, 45%, and 10% of living termite species, respectively.

We do not know what substrates soil feeders really eat. Soil feeders can generally be distinguished from litter feeders by intestinal morphology, the stable isotope ratios of carbon and nitrogen (Fig. 22), and by the higher activity of certain gut bacteria, notably methanogens and organisms able to ferment reduced and recalcitrant substrates, including aromatics. An interesting idea that soil feeders eat microorganisms such as bacteria and fungi in ingested soil, and therefore are important members of microbial chain in the tropical forests, is partly supported by the stable isotope analysis and the detection of lysozyme decomposing cell wall of bacteria, but this theory has not been proved.

B. Functional Role of Termites

Taxonomic composition and feeding habits of termite assemblages in some African habitats are shown in Table VI. The species diversity is the highest in the tropical forest and decreases from forests to arid ecosystems. The dominant feeding habits gradually change from soil-feeding to wood- and litter feeding (cellulose and lignocellulose feeding). The species diversity of fungus growers (Macrotermitinae) is the highest in slightly dry areas such as savannas and dry forests where soil fungi are poorer than in moist forests. A similar tendency is recognized in tropical forests of southeast Asia.

The change may be explained in relation to the activities of fungi. Soil feeders are abundant and diversified in tropical rain forests where fungi, in particular white rot, may be active and decompose lignocellulose efficiently, decreasing the litter accumulation on the ground. On the other hand, lignocellulose feeders cultivating white rot of *Termitomyces* in the nest chambers with high humidity and temperature are dominant in slightly dry tropical areas where the low humidity may suppress the decomposition of lignocellulose by white rot. Cellulose feeders are widely distributed in areas where fungi may be inactive due to low temperature and inadequate humidity.

Abundance of termites sometimes exceeds 10000/m^2 and 10 g wet weight/m^2, although it is usually up to 4500/m^2 and 10 g wet weight/m^2. The biomass of 10 g wet height/m^2 is equivalent to 250 persons/km^2 and 40 kg/person in human beings. In the tropical rain forest of Malaysia where fungus growers are dominant, as much as 30% of the leaf litter supply is consumed by termites. In a Guinea savanna in Nigeria where fungus growers are dominant, termites consume up to 55% of surface litter. Termites are responsible for up to 20% of total carbon mineralization in the savannas of Africa. Consumption of dead plant material by termites increases when fungus growers are dominant, because they consume five to six times more food per unit of biomass than other termites.

Termites are consumed by a great variety of animals ranging from ants to human beings, forming a large detritus chain. Furthermore, a large number of alates produced by mature termite colonies are eaten by birds

TABLE VI

Taxonomic Composition and Diet Structure of Termite Communities from Various African Savannas and Rain Forests

				Number of species eating					Number of species					
Locality	Type of vegetation	Latitude	Annual rainfall (mm)	Grass and grass litter	Fresh wood and leaves litter	Decomposing litter	Humus and soil	Kalotermitidae	Rhinotermitidae	Termitinae	Apicotermitidae	Macrotermitidae	Nasutitermitidae	Total
Fété Olé, Senegal	Wooded steppe (small sand dunes) (reference area)	16°N	375	8	11	0	3	0	3	8	0	3	5	19
Tsavo east, Kenya	Bush and wooded steppe	3.5°S	400	8	17	0	3	4	1	9	0	8	3	25
Cap Vert, Senegal	Woody sub-Guinean savanna	15°N	550	6	17	1	8	1	3	12	2	6	4	28
Mokwa, Nigeria	Primary and secondary woodland	9°N	1175	10	17	1	9	0	1	9	3	11	7	31
Lamto, Ivory Coast	Grassy to woody derived savanna	6°N	1290	10	18	3	13	2	1	10	4	9	10	36
Youhouli, Ivory Coast	Grass savanna	5°N	2000	10	(0)	1	12	0	0	7	5	7	3	22
Edea, Cameroon	Equatorial rain forest	3.5°N	>3000	0	8	4	31	0	1	25	9	5	3	43

From Josens, 1985.

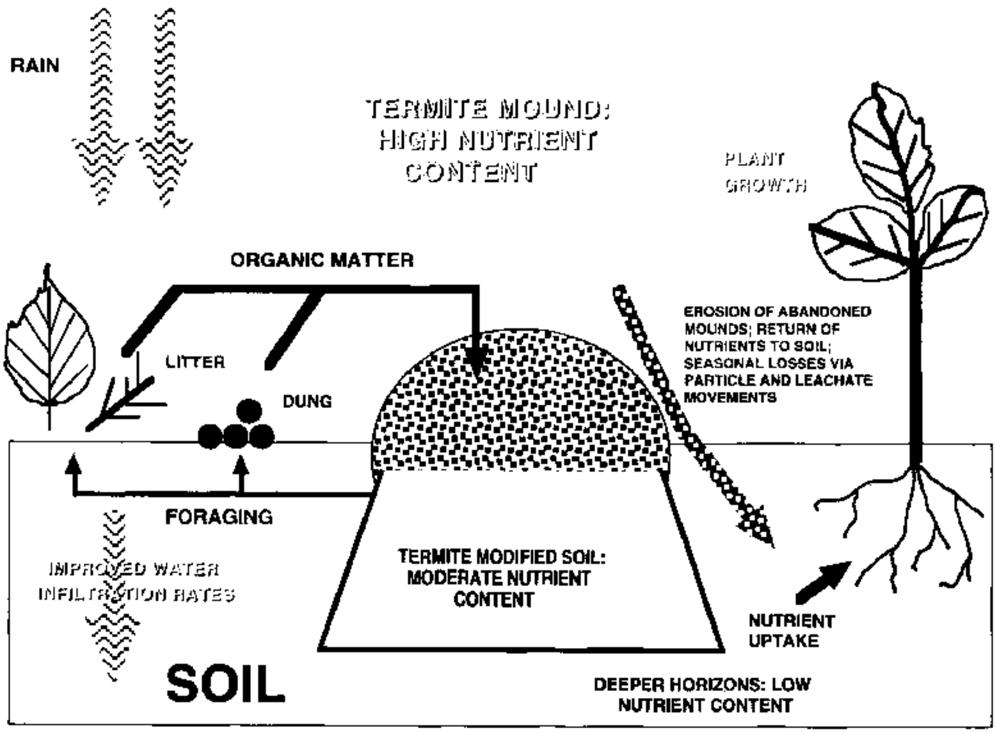


FIGURE 23 Schematic diagram depicting the influence of termites on plant growth. From Conventry *et al.*, 1988.

and spiders in the canopy, joining the grazing chain of terrestrial ecosystems.

C. Termites and Soil

While the distribution and abundance of termites are influenced by climatic conditions (air temperature and humidity), soil properties, and vegetation, termites modify local climate conditions, soil properties, and vegetation through the activities of building mounds and excavating subterranean galleries. The effect of termites on soils has been well examined in the textbook *Termites and Soils* by Lee and Wood (1971).

Termites can modify the soil profile by removing soil from various depths (up to 70 m) and transporting it to the ground surface in the form of runways, sheetings, or mounds, from where it is redistributed by water and wind erosion. In central Africa, termites of *Macroterms* form huge mounds up to 10 m in height and 30 m in diameter. Excavation of subterranean galleries has positive effects on the hydraulic conductivity and infiltration rates.

Large mounds built by species of Macrotermitinae in Africa and Asia often support a vegetation comprising trees and shrubs that is quite distinct from that of the surrounding soils. This difference is considered due to protection from fire, improved drainage, greater soil depth, higher soil moisture, and improved soil fertility. Cation change capacities (CEC), exchangeable cations (Ca, Mg, Na, K), and base saturation decrease with increasing distance from the mound of *Macrotermes michaelseni* in the grassland of Kenya and available phosphorous and NO_3^- nitrogen are high near the mound. This causes the change of vegetation; maximum standing stock of plants occurs near the mound base and the number of plant species increases with increasing distance from a mound. Figure 23 shows a schematic diagram depicting the influence of termite mounds on plant growth.

The termite mound supports a higher level of bacterial activity than adjacent top soil, especially of cellulose decomposers, and the activities of these microorganisms result in the release of nutrients into the mound soil. Many species of higher termites have developed to store the fragments of grasses and leaves in the mounds mainly in somewhat dry regions. The stored fragments are heavily colonized by microorganisms. The storage chambers may have the specific function of decomposition chambers and the process may be similar to the fungus gardens of the Macrotermitinae in that it provides the colony with organic material enriched in nitrogen.

TABLE VII

Termite Abundance and Biomass in Tropical Regions (AFR, African; NEO, Neotropical; ORI, Oriental

Ecosystems	Region	Abundance nos. m^{-2}	Biomass g m^{-1}
Primary or near-primary forest			
Riverine forest (Congo)	AFR	1000	ca. 11
Riverine forest (Nigeria)	AFR	2646	6.9
Semideciduous forest (Nigeria)	AFR	3163	8
Semideciduous forest (Cameroon)	AFR	2282–6957	8.31–123.2
Rainforest (Puerto Rico)	NEO	87–104	0–1
Rainforest (Amazonia)	NEO	1865	6.0–7.5
Rainforest (Pasoh, W. Malaysia)	ORI	3200–3800	ca.9
Wet evergreen (S. India)	ORI	2011–2449	3.62–3.72
Montane (Gunung Mulu, E. Malaysia)	ORI	38–295	0.01–0.7
Rainforest (dipterocarp, E. Malaysia)	ORI	779–1603	1.1–1.8
Rainforest (kerangas heath forest)	ORI	2271	ca.4
Secondary forest or tree plantation			
Terminalia plantation (1 yr) (Cameroon)	AFR	244–2959	0.43–19.89
Terminalia plantation (5 yr) (Cameroon)	AFR	5170–6703	10.98–35.13
Secondary forest (40 yr) (Cameroon)	AFR	2328–10488	39.11–114.16
Tropical savanna			
Grass savanna (Central Africa)	AFR	612–701	1.3–1.9
Derived savanna (W. Africa)	AFR	861	1.7
Guinea savanna (W. Africa)	AFR	2966	3.6
Guinea savanna (W. Africa)	AFR	4402	11.1
Acacia savanna (S. Africa)	AFR	732	0.98
Broadleaved parkland (South Africa)	AFR	3300	8.43
Agro-ecosystems			
Grazed pasture (W. Africa)	AFR	2010	2.8
Maize (1 yr, W. Africa)	AFR	1553	1.7
Maize (8–24 yr, W. Africa)	AFR	6825	18.9
Sugar cane (W. Africa)	AFR	4800	5.4
Temperate scrubs			
Mediterranean scrubs (W. Australia)	AUS	27–45	0.97–1.93

D. Termites and Human Disturbances

The abundance and species diversity of termites tend to decrease when tropical forests and savannas are cleared and modified (Table VII). At first soil feeders are replaced by fungus growers and then by wood feeders in Africa and Asia. However, the destruction of natural ecosystems and their replacement by other ecosystems in some cases provides suitable habitats for some groups of termites, leading to an increase in their population. Zimmerman *et al.* (1982) were the first to draw attention to the topic in relation to the production of greenhouse gas by termites.

The global concentration of atmospheric methane, a greenhouse gas, is increasing in recent years at an annual rate of about 1%. This has been mainly attributed to paddy rice cultivation, enteric fermentation (mainly by ruminants), biomass burning, and fossil fuel consumption. Termites have high biomass and emit methane in many tropical ecosystems.

Much research has been done on the global estimates of termites, methane production by various groups of termites, oxidation of methane produced by termites in the soil, and the importance of methane production by termites on global climatic change. The rate of emission of methane by termites differs between species and feeding habits. Soil feeders tend to produce more methane than wood feeders on a biomass-specific basis. An important fact is that a large amount of methane produced by subterranean termites is oxidized by microorganisms in the soil. Thus the annual contribution by termites is estimated to be less than 20 Tg and probably less than 10 Tg (ca. 4% and 2% of global total from all sources, respectively).

Termites usually consume dead plant material and do not attack living plants in natural forests. However, some termites that are able to survive in agricultural ecosystems become harmful to crops (Harris, 1971). This tendency is stranger when the crops are exotic. A few hundred species of termites cause damage and only about 50 species are serious pests (Pearce, 1997).

See Also the Following Articles

AMAZON ECOSYSTEMS • INSECTS, OVERVIEW

Bibliography

Abe. T. (1987). Evolution of life types in termites. In *Evolution and Coadaptation in Biotic communities* (S. Kawano, J. H. Connell and T. Hidaka, Eds.), 125–148. University of Tokyo Press, Tokyo.

Abe. T., Higashi, M., and Bignell, D. E. (Eds.) (2000). Termites: Their symbiosis, sociality and global diversification. In press.

Behnke, F. L. (1977). *A Natural History of Termites.* Charles Scribner's Sons, New York.

Higashi, M., and Abe, T. (1997). Global diversification of termites driven by the evolution of symbiosis and sociality. In *Biodiversity* (T. Abe, S. A. Levin, and M. Higashi, Eds.), 83–112. Springer.

Higashi, M., Abe, T., and Burns, T. P. (1992). Carbon-nitrogen balance and termite ecology. *Proc. R. Soc. London B.* 249, 303–308.

Krishna, K., and Weesner, F. M. (Eds.) (1969). *Biology of Termites, Vol. I*, 598. Academic Press, New York.

Krishna, K., and Weesner, F. M. (Eds.) (1970). *Biology of Termites, Vol. II*, 643. Academic Press, New York.

Lee, K. E., and Wood, T. G. (1971). *Termites and Soils.* Academic Press, New York & London.

Wilson, E. O. (1971). *The Insect Societies.* Harvard University Press, Cambridge.

Wood, T. G. (1978). Food and feeding habits of termites. *Production Ecology of Ants and Termites* (M. V. Brian, Ed.), 55–80. Cambridge University Press, Cambridge.

Wood, T. G., and Sands, W. A. (1978). The role of termites in ecosystems. *Production Ecology of Ants and Termites* (M. V. Brian, Ed.), 245–292. Cambridge University Press. Cambridge.

KEYSTONE SPECIES

Bruce A. Menge and Tess L. Freidenburg
Oregon State University

GLOSSARY

dominant species Species that owe their influence to their high abundance. Such organisms account for most of the biomass in a community, and thus are the primary components of community structure. Trees in forests, mussels in rocky intertidal habitats, grasses in grasslands, and kelps or corals in nearshore subtidal habitats are all dominant species.

interaction webs (functional webs) Subset of species that through their interactions and responses to abiotic factors make up the dynamic core of food webs or communities. These webs include keystone species, dominants, and other strong interactors.

key-industry species Prey of intermediate trophic status that support a large group of consumers.

keystone species Consumers that have a large effect, and one that is disproportionately large relative to their abundance, on communities and ecosystems. Uniquely, the strong effects of keystone species on their interacting species exert extensive influence, often indirectly, on the structure and dynamics of communities and ecosystems. They are a distinct subset of a more broadly defined set of "strong interactors" that also include species having strong effects on interacting populations but not necessarily on communities or ecosystems. Keystone species can include predators, parasites, pathogens, herbivores, pollinators, and mutualists of higher trophic status, but generally are not plants, sessile animals, or "resources."

strong interactors (foundation species) Species that have a large effect on the species (one or a few) with which they interact. Communities and ecosystems may have many strong interactors, and such species may occur at all trophic levels. "Strong interactors" is a more general term that can include keystone species, but not all strong interactors are keystone species. A similar idea is "foundation species," defined as the group of critical species whose effects and interactions define much of the structure of a community.

weak interactors Species that have little effect on other species, at least under average conditions. Under some circumstances, weak interactors may occupy important roles in ecological communities as a result of changes that lead to temporary increases in their abundance, size, or biomass.

KEYSTONE SPECIES WERE originally defined as species high in the food web that greatly modify the composition and physical appearance of an ecological community. The original keystone species was a carnivorous sea star occurring in a marine rocky intertidal habitat dominated by macroalgae, mussels, barnacles, and other invertebrates (Paine, 1966). Implicit in the initial definition was the idea that keystone species have effects that are disproportional to their abundance. This concept was formalized by Robert Paine in the 1960s: "the patterns of species occurrence, distribution and density are disproportionately affected by the activities of a single species of high trophic status" (see Power *et al.*, 1996). With time, the number of examples of apparent keystone species grew, leading to heightened awareness that the phenomenon was more general, and therefore more important, than originally thought. This increasing familiarity with, and embracing of, the concept of keystone species led workers to broaden its definition and, perhaps inevitably, to extend it to questionable applications (Mills *et al.*, 1993). Thus, in addition to keystone predators, workers defined "keystone herbivores," "keystone prey," "keystone mutualists," "keystone hosts," "keystone resources," "keystone guilds," and "keystone modifiers." One idea, the "extended keystone hypothesis," considered that every community or ecosystem has several keystone species, which are those that dominate community structure and dynamics. Simultaneously, many researchers simplified the definition of keystone species to refer to those members of a community that had some large effect, regardless of the keystone's relative abundance or trophic status. These alterations led Mills and coworkers to question the usefulness of the concept. Their critique stimulated a comprehensive review of these and related issues (Power *et al.*, 1996; see Section II). One outcome of the recent controversy has been to refocus attention on keystone species and the overall importance of this concept in understanding the dynamics of ecological communities. This article surveys the roots of the issue, considers the precise meaning of the term, and reviews several case studies. It then addresses problems in methods of documenting keystone effects, in the identification of keystone species, and in determining the context under which keystone effects are likely. Finally, it evaluates the concept in the broader context of the strength of interactions among species, and concludes with comments on future directions of research.

I. HISTORICAL PERSPECTIVE

A. Identification of *Pisaster ochraceus* as a Keystone Species

In 1966, Paine (1966) reported the results of a study of the role of predatory sea stars in structuring marine intertidal communities on a rocky shore on the outer coast of Washington State. In temperate regions, such communities typically display a striking spatial pattern called "zonation." The hallmark of this pattern is that dominant space-occupying organisms are arranged in vertically stacked horizontal bands. For example, temperate shores commonly have a band of barnacles and fucoid algae on the upper shore, a band of mussels on the middle shore, and a band of macroalgae on the lower shore. On the Washington coast, the macroalga-dominated "low zone" also harbors various mobile and sessile invertebrates, including limpets, chitons, sea urchins, whelks, anemones, and, significantly, the sea star *Pisaster ochraceus*. A particularly intriguing feature of the zonation pattern was the sharp demarcation between the mussel bed and the low algal zone. Observing that low-zone-occupying sea stars fed regularly on mussels, and that the only abundant source of mussels was higher on the shore in the middle zone, Paine postulated that predation by sea stars determined the sharp lower limit to mussels. Specifically, he speculated that at high tide, sea stars moved up from their low tide resting places in the low zone to remove prey from the lower edge of the mussel bed, then retreated back to the low zone where the prey were consumed.

To test the hypothesis that sea star predation determined the lower limit of the mussels, he periodically removed *Pisaster* from a section of a rocky outcropping and compared this section to an area in which sea stars had been left at natural densities. Within 3 years the mussel *Mytilus californianus* had increased in abundance, causing a reduction in local species richness of macroorganisms from 15 to 8 species (Paine, 1966). Ultimately, after 10 years of excluding sea stars, a single-species monoculture of *M. californianus* dominated the shore (Fig. 1). Ancillary studies suggested that this change depended on two important mechanisms. First, *Pisaster* preferentially fed on mussels. This sea star actually will feed on many invertebrate species, but given a choice, it selects mussels. Second, mussels were dominant competitors for space. Field observations, and later experiments, indicated that mussels could displace all other occupants of space, including macrophytes, sessile invertebrates, and mobile invertebrates, by crowding them out. In so doing, mussels also created habitat for a sharply different set of invertebrate cryptofauna that occupied the byssal "forest" (the fibers that mussels use to attach to the rock) beneath the mussels.

Paine later repeated this experiment on a nearby coastal island (Tatoosh Island) with essentially identical results, further demonstrating that under natural conditions *Pisaster* indirectly enhances the persistence of other space occupiers by preventing mussels from in-

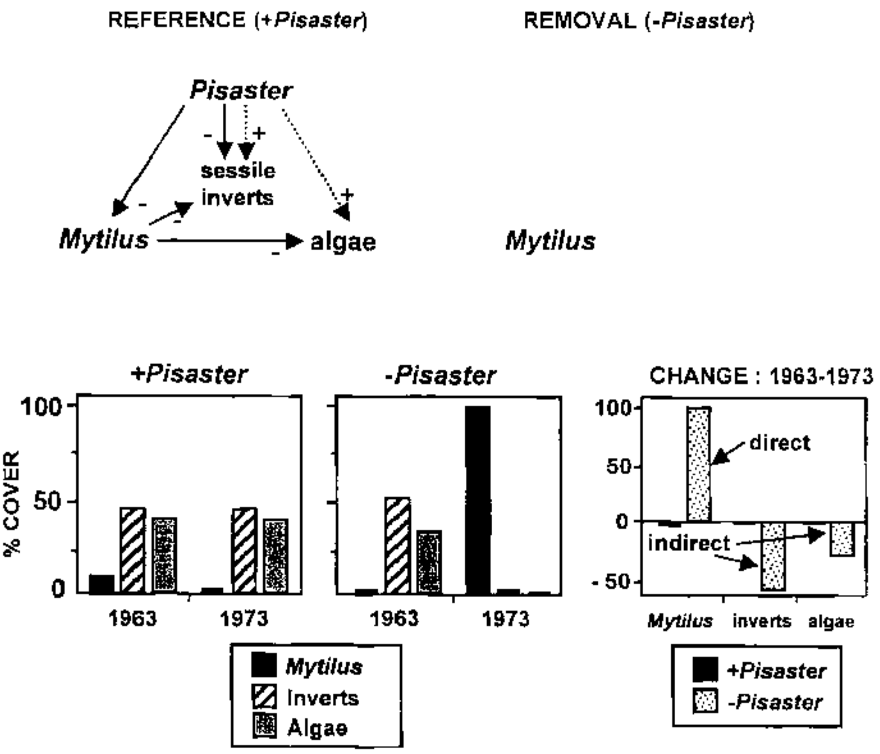


FIGURE 1 Results of *Pisaster* removal experiments. Upper diagrams: "Reference" shows the natural interaction web in the low zone, and "Removal" shows the interaction "web" (i.e., *Mytilus*) 10 years after removal began. Arrows point to the species or group affected. Solid arrows show direct effects; dotted arrows show indirect effects. "Minus" by an arrow indicates that the effect was negative; "plus" by an arrow indicates that the effect was positive. Lower diagrams: left and center show covers of *Mytilus*, other sessile invertebrates (barnacles, anemones), and algae in 1963 (beginning of experiment) and 1973 in the presence and absence of *Pisaster*. Right diagram shows difference between covers in 1973 and 1963 for each group of sessile organisms. "Direct" indicates that increase in mussel cover was due to direct predation by sea stars. "Indirect" indicates that decreases in sessile invertebrates and algae were an indirect consequence of sea star predation, through competitive elimination of sessile invertebrates and algae by mussels. (Data from Paine, 1974, cited in Power *et al.*, 1996.)

vading the low intertidal zone. The striking difference in species composition and physical appearance of the low zone community wrought by the sea star, a community member that was nowhere close to being "dominant" either numerically or in biomass, was the basis for Paine's identification of *Pisaster* as a keystone predator. Note that contrary to some uses of the concept, the original definitions did not comment on keystone species as determinants of species diversity. Although the changes in species diversity that occurred in Paine's (1966) field experiments revealed an important role of such species, Paine and other early workers on the topic (Estes *et al.*, 1978) evidently did not consider these changes to be fundamental to the concept of keystone species.

The formulation of the keystone species concept contributed to a gradual shift in ecological thought. Throughout the 1960s and 1970s, most ecologists believed that competition was the most important process structuring communities. By the late 1960s, advances in ecological theory and modeling, in conjunction with comparative and experimental field studies, demonstrated that processes other than competition could be important. Experiments such as those leading to the keystone species concept were among the first to describe how predation can dramatically affect community structure. Following Paine's pioneering work, ecologists began to identify keystones in other systems.

B. Are There Others? Keystone Predators Elsewhere

One of the first documented examples of a keystone predator in another system was the sea otter, *Enhydra*

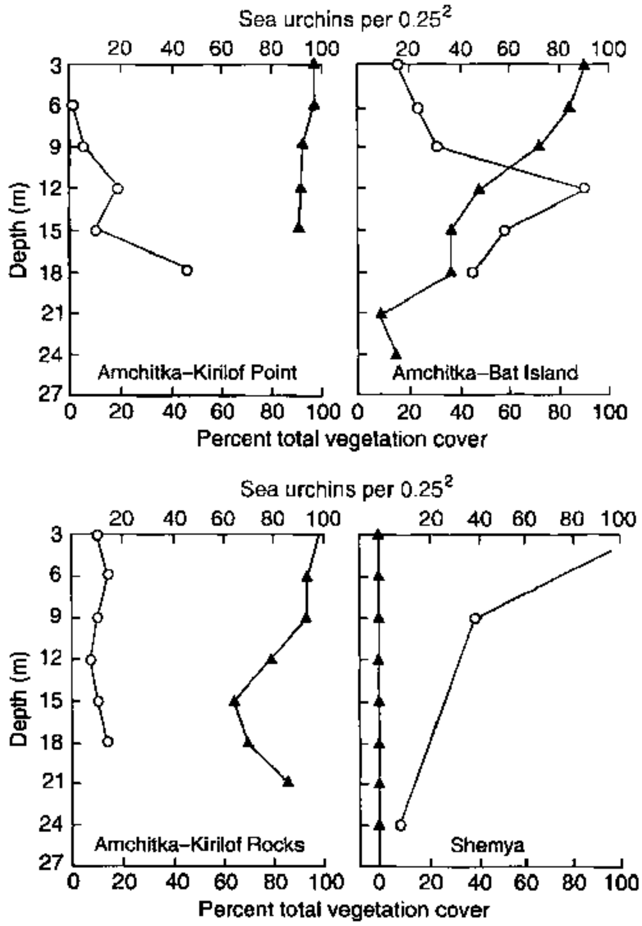


FIGURE 2 Total cover of macroalgae and sea urchin density by depth at three locations on Amchitka Island—Kirilof Point, Bat Island, and Kirilof Rocks (all with sea otters)—and one at Shemya Island (without sea otters). ▲, vegetation cover; ●, sea urchin density. (From Estes *et al.*, 1978.)

lutris. Estes and coworkers (Estes *et al.*, 1978) observed that subtidal communities in the Aleutian Islands, Alaska, where sea otters were abundant, differed from those in nearby areas where sea otters had been locally extirpated by hunting. In waters surrounding islands with sea otters, kelp forests at shallow depths dominated nearshore subtidal communities, and within these kelp forests sea urchins, an important consumer of kelp, were relatively small and urchin biomass was low (Fig. 2). On neighboring islands without sea otters, however, sea urchins were large, urchin biomass was high, and kelps were absent. On the basis of these data and the observation that sea otters preferentially feed on sea urchins, Estes and coworkers postulated that sea otters are a keystone species in this community. They argued that predation by sea otters reduced sea urchin grazing, thereby maintaining kelp forests and associated species. By documenting the "keystone predator" phenomenon in a different ecosystem, this research bolstered the view that such dynamics were potentially widespread.

Subsequent work both supported this hypothesis and expanded our understanding of the conditions facilitating a keystone role for sea otters. In a large-scale study published in 1995, Estes and his colleague David Duggins determined the spatial generality of sea otter effects and tested the prediction that after sea otters colonize new areas, sea-urchin-dominated subtidal communities would become kelp-dominated. Surveys in the Aleutians and in southeast Alaska showed that the differences associated with sea otter presence or absence in the earlier studies were general in space—kelps were abundant in the presence of sea otters and scarce in the absence of sea otters. Further, considering all surveyed sites, sea urchin and kelp abundance were strongly inversely correlated (Fig. 3). Finally, sea urchin abundance declined sharply at sites invaded by sea otters but did not change at sites where sea otter abundance remained constant. An interesting regional difference was that rates of increase in kelp abundance in the presence of invading sea otters were higher in southeast Alaska. Estes and Duggins suggested that higher rates of kelp increase resulted from differences in the recruitment of sea urchins, which was much greater in the Aleutians than in southeast Alaska. Coupled with sea otter preference for larger, adult urchins, higher recruitment rates of juvenile urchins presumably led to moderate rates of grazing on kelp sporelings and juveniles in the Aleutians, slowing rates of increase in kelp abundance relative to that in southeast Alaska.

A recent complication with troubling conservation implications suggests that killer whale (*Orcinus orca*) predation has begun to eliminate sea otters from their native habitats with predictable consequences for abundances of sea urchins and kelp (Fig. 4). During the 1990s, sea otter abundance declined sharply at several Aleutian Islands. Once freed from sea otter predation, sea urchin populations increased, leading to severe declines in kelp and associated species. Estes and coworkers suggested that killer whales may have shifted their diet from large marine mammals to sea otters ultimately as a consequence of human activity. Overfishing and increased ocean temperatures (due to global climate change) are associated with declines in forage fish for sea lions and seals, and consequently these marine mammals have declined sharply in number. Historically these large marine mammals have been the primary prey of killer whales; in their absence, some orcas have shifted their diet to otters. Because sea otters are small

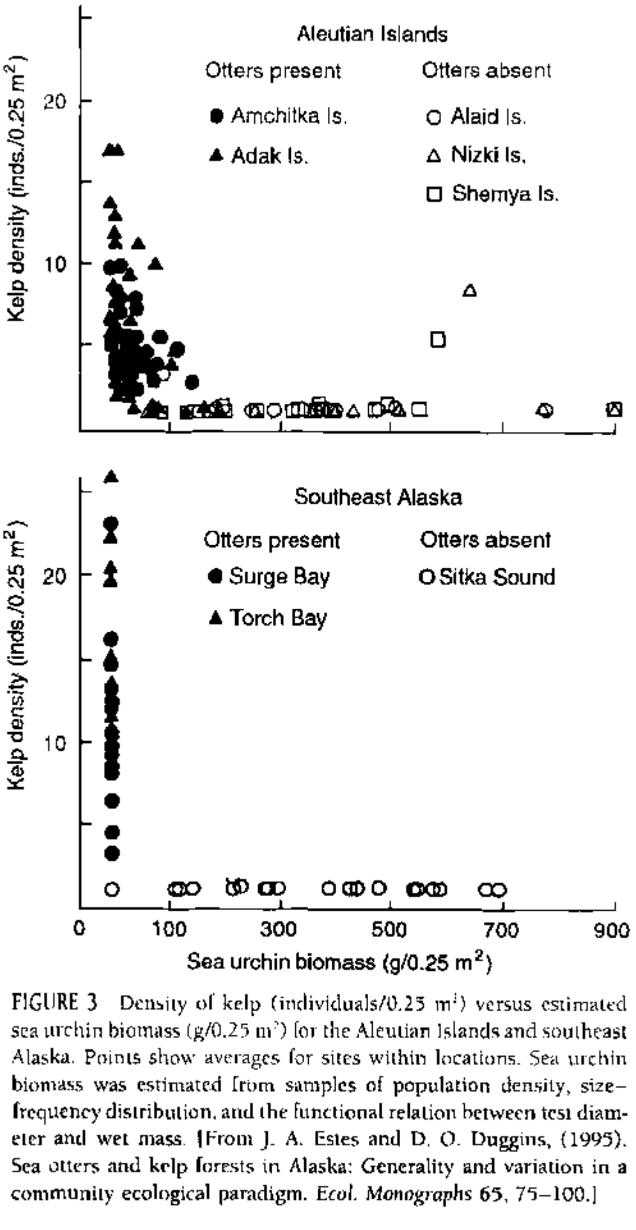


FIGURE 3 Density of kelp (individuals/0.25 m^2) versus estimated sea urchin biomass (g/0.25 m^2) for the Aleutian Islands and southeast Alaska. Points show averages for sites within locations. Sea urchin biomass was estimated from samples of population density, size-frequency distribution, and the functional relation between test diameter and wet mass. [From J. A. Estes and D. O. Duggins, (1995). Sea otters and kelp forests in Alaska: Generality and variation in a community ecological paradigm. *Ecol. Monographs* 65, 75–100.]

relative to pinnipeds, rates of predation loss have evidently been high, with substantial direct and indirect consequences for subtidal communities.

C. Generalization and Liberalization of the Definition: Other Types of Keystone Species

These earliest examples of keystone species were of consumers that modified the community through predation. However, some investigators suggested that species other than top predators could play keystone roles in communities. Subsequently, a variety of different types of organisms were termed "keystone" species, including herbivores, plants, pollinators, pathogens, habitat modifiers, and mutualists. The remainder of this article presents examples of such alternative keystones, summarizes a critique of these ideas, and offers proposals for further clarification of keystone and other functionally important species.

1. Herbivores

By consuming primary producers, keystone herbivores can have dramatic impacts on community structure. Among species identified as keystone herbivores are kangaroo rats and bison. In the Chihuahuan Desert, Brown and colleagues (see summary in Brown, 1998) suggested that a guild of kangaroo rats played a keystone role. In plots where kangaroo rats were excluded, grass cover increased threefold. This was accompanied by a change in species composition; species shifted from those typical of desert shrubland to those characteristic of arid grasslands (Fig. 5). The mechanisms responsible for this transition were seed predation and soil disturbance by kangaroo rats. By preferentially eating the seeds of competitively dominant grasses, kangaroo rats indirectly released subordinate plant species from competition. Furthermore, kangaroo rat burrowing favored disturbance-tolerant plant species. Hence, the presence or absence of kangaroo rats determined whether the community was a desert shrubland or an arid grassland.

Note that this example both extends the concept to herbivores and attributes the keystone effect to a multispecies group instead of a single species. Recently Brown (1998) suggested that the kangaroo rat effect was attributable primarily to two species, *Dipodomys merriami* and *D. spectabilis*, and one of these (*D. spectabilis*) went locally extinct in 1994. If the system persists unchanged well beyond this extinction, these fortuitous changes may suggest that *D. merriami* is a keystone species. In our view, however, the original interpretation of these herbivores as a keystone "guild" may contribute to confusion in terminology. This study is an excellent example of the strong effects of consumers on community structure, but as explained later, it was not designed to discern between keystone versus "diffuse" (multispecies) predation.

Bison have also been identified as keystone herbivores (Knapp *et al.*, 1999). Studies at Konza Prairie have documented strong effects of bison on species composition, diversity, and several physical and chemical aspects of ecosystem function of grassland commu-

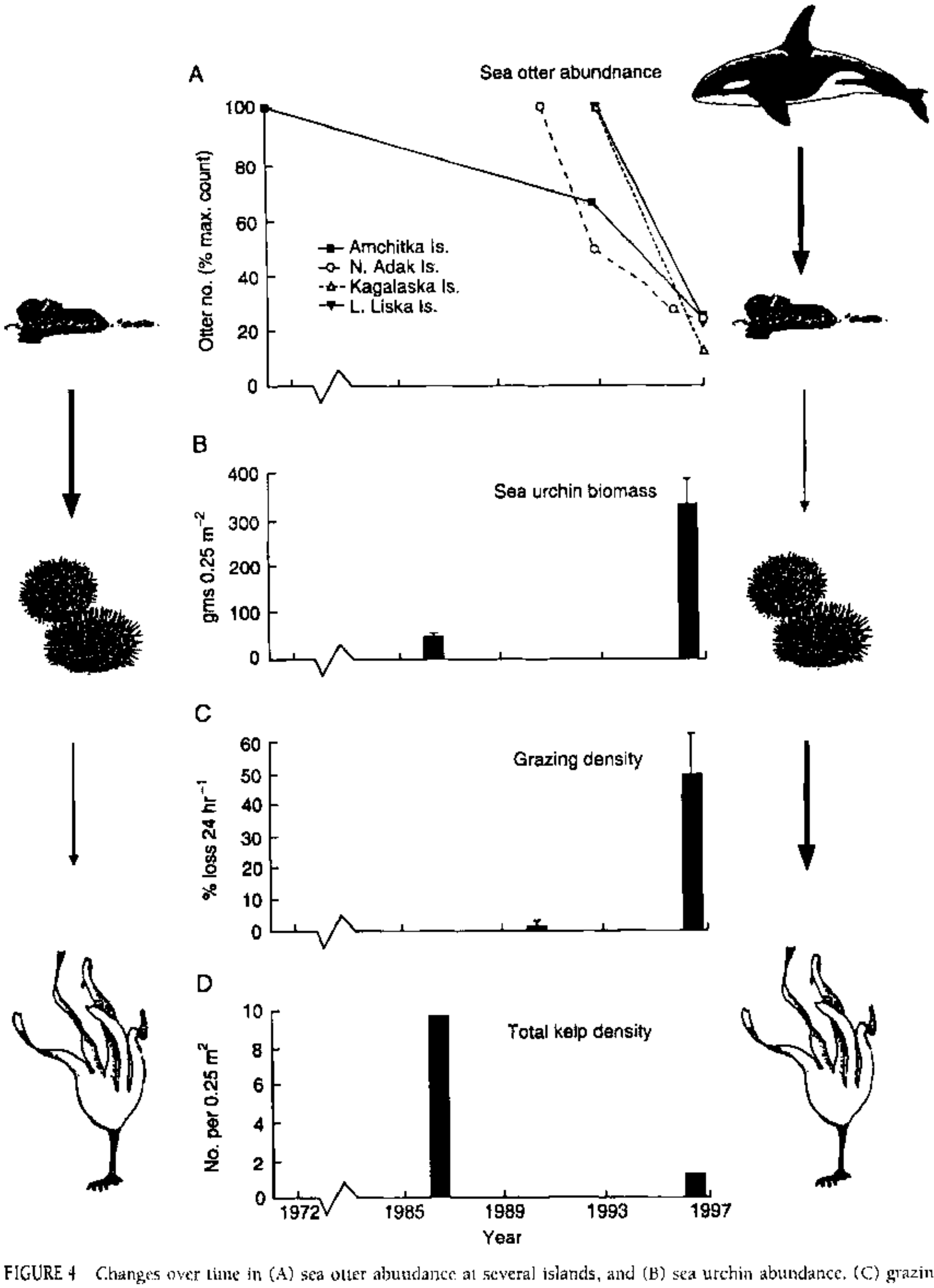


FIGURE 4 Changes over time in (A) sea otter abundance at several islands, and (B) sea urchin abundance, (C) grazing intensity, and (D) kelp density at Adak Island in the Aleutian archipelago. Error bars are 1 SE. Diagrams to the left and right of the data panels suggest the mechanisms leading to the changes. The left diagram shows kelp forest dynamics without orcas and the right diagram shows changes induced by the addition of orcas as the top predator. Thick arrows show strong effects, and thin arrows show weak effects. [From J. A. Estes *et al.* (1998). Killer whale predation on sea otters linking oceanic and nearshore ecosystems. *Science* 282, 473–476. Copyright 1998 American Association for the Advancement of Science.]

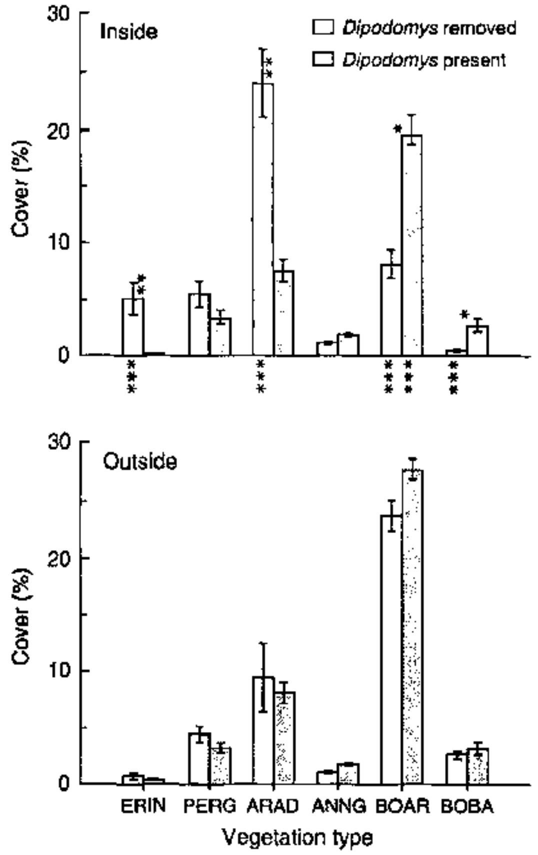


FIGURE 5 Densities of perennial and annual grasses in the presence and absence of kangaroo rats (*Dipodomys*). Data are mean percent cover (±1 SE) of *Eragrostis lehmanniana* (ERIN), other tall perennial grasses (PERG), *Aristida adscensionis* (ARAD), other tall annual grasses (ANNG), and the short annual grasses *Bouteloua aristidoides* (BOAR) and *B. barbata* (BOBA). Top: Data from transects inside enclosures from which kangaroo rats had been removed or allowed to remain. Asterisks above bars indicate significant differences between +kangaroo rat and −kangaroo rat enclosures. Bottom: Data from transects outside the respective enclosures. Asterisks between panels indicate significant differences between transects inside and outside plots where kangaroo rats were present or absent. Analyses employed ANOVA with plots as units of replication; *, $p < 0.05$; **, $p < 0.01$; ***, $p < 0.001$. [From J. H. Brown and E. J. Heske (1990). Control of a desert-grassland transition by a keystone rodent guild. *Science* 250, 1705–1707. Copyright 1990 American Association for the Advancement of Science.]

nities. These patterns occur across scales ranging from the plant level to patches to landscapes, and are largely a consequence of bison preference for competitively dominant grasses and for persistent habitat features. As will be discussed later, Knapp and coworkers also evaluated the question of whether or not bison are truly keystone herbivores using the operational procedure recommended by Power *et al.* (1996).

2. Plants (Resources)

Terborgh (1986; cited in Power *et al.*, 1996) suggested that certain plants in Peruvian rain forests can provide keystone "resources." Most tree species in these tropical rain forests fruit in synchrony. However, when these species are out of season, palm nuts, fig trees, and nectar-bearing plants can provide a crucial resource to frugivores. Less than 1% of the plant biomass of tropical forests is made up by these species and yet virtually all frugivorous animals rely on them during the 3 months when other food sources are rare. Consequently, Terborgh argued that these plants maintained high animal diversity in these communities and therefore occupied a keystone role. Although Terborgh's work focused on Peru, he suggested that keystone plant resources may be widespread in tropical rain forests.

In another example, anadromous fishes swimming upstream to spawn, or their carcasses, were suggested to be keystone resources. It was argued that a large number of vertebrate predators and scavengers, especially terrestrial species, rely on the energy input provided by anadromous fishes returning to their natal streams. Such energy sources can be strong interecosystem links and are suggested to have an influence on the diversity of terrestrial mammal assemblages.

We do not dispute the notion that such species or resources can be important to community dynamics, but for two reasons we question calling them keystones. First, in most such cases, the evidence for keystone effects is weak and conjectural; the investigators have not demonstrated that removal of the resources would lead to wholesale community changes. Second, with many others we believe that top-down effects (such as keystone predation, keystone herbivory) are qualitatively distinct from bottom-up effects (e.g., primary production, food supply). We therefore suggest that examples in which plant-level processes are thought to determine consumer effects, and which thus suggest that the plant is the keystone species, serve to confuse rather than clarify the concept by combining under a single idea a heterogeneous collection of ecological processes. Although there is increasing evidence in favor of the role of bottom-up effects as important determinants of community dynamics, these effects commonly lead to variation in top-down effects, often with single species (i.e., keystones) having large effects that are disproportionate to their abundance. We therefore argue that retaining the original distinction of keystone species as consumers is simpler and clearer, and hence better serves both understanding and insight into community dynamics and policy decisions regarding ecosystem management and conservation. Studying the role of bottom-up processes as determinants of community structure is an exceedingly important area of research, but we believe progress both in this area and in the role of keystone species will be enhanced if they are treated as distinct but dynamically linked concepts.

3. Mutualists (Pollinators, Seed Dispersers)

Flying foxes (*Pteropus* spp.) have been proposed to be keystone "mutualists" in tropical rain forests. Flying foxes are large tropical bats that are important pollinators and seed dispersers in Old World forests and on tropical islands. Especially on isolated islands where other vertebrate pollinators are scarce, these bats can be responsible for dispersing 80–100% of the seeds. Cox and coworkers (cited in Power *et al.*, 1996) argued that, through these pollination and dispersal "services," flying foxes may be responsible for maintaining high plant diversity in the forest communities in which they occur. Comparisons of the proportion of fruits and seeds dispersed by flying foxes on Guam, where flying foxes have been driven nearly to extinction, and on Samoa, with abundant flying foxes, were consistent with this idea (see Rainey *et al.*, 1995, cited in Power *et al.*, 1996). Between 0 and 100% of fruits were dispersed by flying foxes on Western Samoa, whereas 0–1% of seeds were dispersed by flying foxes on Guam. Such data are only suggestive, however; further study is needed to demonstrate that forest community structure reflects these differences in fruit and seed dispersal, and that flying foxes play a keystone role in determining community structure.

4. Habitat Modifiers

Organisms that influence the availability of resources for other species by modifying the physical environment also have been considered keystone species (more recently the term "ecosystem engineers" has been coined to identify such organisms; see Lawton and Jones, 1995, cited in Power *et al.*, 1996). Beavers, red-naped sapsuckers, kangaroo rats, and prairie dogs are all examples of species that, through their nonforaging activities, modify and/or create habitat for other species. By altering the hydrology of rivers, beaver dams can have profound effects on sediment retention, nutrient cycling, and the condition of the riparian zone (Naiman *et al.*, 1986). Sapsuckers create habitat for two species of swallows by drilling holes in aspens. In addition, their feeding holes in willows create sap flows, providing a resource for several species of birds, mammals, and insects (see Daily *et al.*, 1993, cited in Power *et al.*, 1996). The activities of kangaroo rats and prairie dogs cause soil disturbance, which affects community structure. These habitat modifiers were all thought to have impacts disproportionate to their abundance and were thus dubbed keystone species.

II. CRITIQUE AND REEVALUATION

Paine's original definition of keystone species referred to top predators that greatly modified the species composition and physical appearance of the ecosystem by preferentially feeding on dominant competitors for space. Although not quantitative, the term was clear in concept and interpretation. As documented in the previous section, however, the term was subsequently applied to a variety of organisms other than top predators and to actions that affected communities in ways other than feeding on competitive dominants. Application to groups of species (e.g., keystone guilds) rather than to a single species blurred the concept even further. These definitional liberalizations, combined with a lack of rigor in determining if candidate species meet the necessary criteria to be termed "keystone," eventually undermined the usefulness of a potentially powerful concept (Mills *et al.*, 1993). Here we evaluate these issues and suggest a scheme of classification and definition (slightly modified from that offered by Power *et al.*, 1996) that we believe further clarifies and advances the usefulness and application of the concept of keystone species.

A. Reevaluating the Keystone Species Concept

The seeming lack of a working definition of keystone species prompted Mills *et al.* (1993) to argue that careless usage had sufficiently degraded the value of the concept to justify removing it from usage. They further argued that because the term was poorly defined it was functionally useless and conservation strategies therefore should not be based on protecting keystone species. They advocated focusing on interaction strength rather than on a species' keystone or nonkeystone status as a more useful management strategy.

In response to the criticism by Mills *et al.* (1993), a group of ecologists with expertise in the study of keystone species and strongly interacting species met to evaluate the keystone species concept. The publication that resulted from this meeting was a signal achievement, and much of the present article is patterned after this synthesis (Power *et al.*, 1996). The group agreed with Mills *et al.* (1993) that through misapplication and questionable redefinition, ecologists and conservation biologists had obscured the meaning of the term keystone species. Rather than abandoning the concept, however, the group proposed clarification and adherence to a set of more sharply defined concepts for community dynamics. They defined a keystone species as "one whose impact on its community or ecosystem is large, and disproportionately large relative to its abundance" (Power *et al.*, 1996). This definition retained the essence of Paine's original usage, but expanded it more broadly to include species other than predators.

Power *et al.* (1996) stressed the importance of having a rigorous, quantitative method of assessing the community- or ecosystem-level effects of a species when determining if it is a keystone. To assess a species' impact on a community they proposed a community importance (CI) index, which in practical terms is quantified experimentally as

$$CI_i = [(t_N - t_D)/t_N](1/p_i)$$

where p_i is the proportional abundance of the species i before it was deleted, t_N (for "normal") is a quantitative measure of a community or ecosystem trait under usual conditions (e.g., productivity, nutrient cycling, species richness, relative abundances of species), and t_D (for "deleted") is the trait in the absence of species i. Keystone species are those whose community importance (CI) is large relative to that of other species (Fig. 6). On the basis of these considerations, Power *et al.* (1996) suggested that if the total impact of a species is plotted against its proportional abundance, keystone species would cluster toward the upper left of the graph in Fig. 7. Power *et al.* (1996) defined another group of species, "dominants," as those that had large total impacts but in proportion to their abundance—these would lie to the upper right of the abundance–impact diagram.

Importantly, keystone species are not necessarily simply "strong" interactors, at least as we define the terms. Keystone species not only have disproportionately large effects, but also have a community-wide or ecosystem-wide impact through direct and indirect effects cascading through the system (Power *et al.*, 1996; see Fig. 1). Strong interactors are species that have a large impact on the species with which they interact, but this large impact could affect only a single interacting species. Only when strong interactors have multispecies effects that alter the structure of the community, or the functioning of the ecosystem, are they also keystone species.

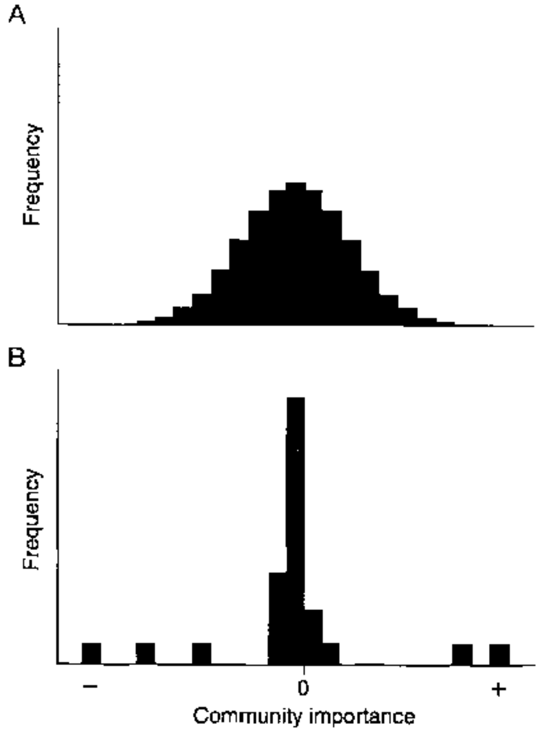


FIGURE 6 Possible frequency distributions of community importance (CI) values for all species in a community. Positive values occur when a community trait decreases after a species is deleted. For instance, in the absence of a mutualist, the target dominant species would also decrease. Negative values occur when a community trait increases after removal of a species. For instance, in the absence of a consumer, the target dominant species would increase. Community importance values may be normally distributed around zero (a), indicating that most species have small effects and keystones are rare. In some communities (b), CI values may have several modes, with keystone species indicated by values far from zero. (From Power *et al.*, 1996.)

B. Related Concepts

The idea that not all species have equal significance in community dynamics is an old one, and many efforts have been made to assign names or terms to distinguish species or groups of species with important roles. Elton (1927) advanced the concept of "key-industry species" as single species of animals supporting a large number of consumers (e.g., copepods, herring, anchovies, sea pens). "Foundation species" were defined as the "group of critical species which define much of the structure of a community" (Dayton, 1972). Related ideas include "interaction webs" and "functional webs," which are defined as the subset of strongly interacting species that regulates community structure. A similar idea, the "extended keystone hypothesis," states that ecosystems are controlled by a small set of key plants, animals,

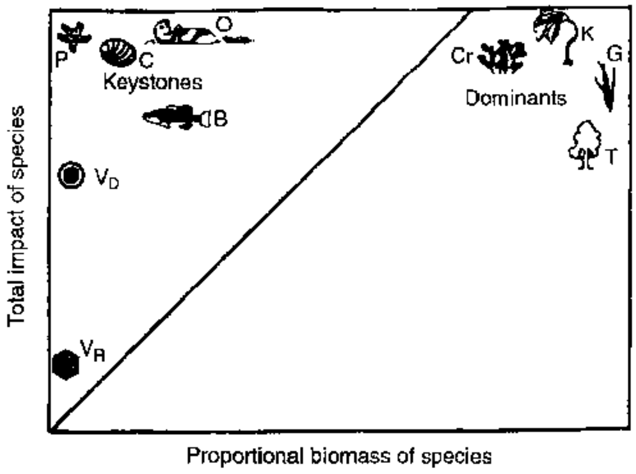


FIGURE 7 Total impact of a species (absolute value of community impact × proportional abundance of a species) versus its proportional abundance, p_i. Species whose total impact is proportional to their abundance would fall along the diagonal line $X = Y$. Keystone species have effects that greatly exceed their proportional abundance, both on a per-capita (or per-biomass) basis and a per-population basis, and would cluster toward the upper left region of the diagram. To illustrate, some organisms (e.g., rhinovirus, V_R, which causes colds in wildlife) may have effects that are greater than expected from their biomass, but because the impact on the community is relatively minor, they are not keystone species. Others (e.g., distemper virus, V_D, a killer of lions or wild dogs) may have collective effects on the community that are disproportionately large, and would be keystone species. Examples of keystone species are *Pisaster* (P), sea otters (O), predatory whelks (*Concholepas*, C), and freshwater bass (B). Dominants are high proportional biomass species whose large effects are not disproportional to their biomass, such as trees (T), giant kelp (K), prairie grasses (G), and reef-building corals (Cr). Positions of each species or group represent educated guesses. (From Power *et al.*, 1996.)

and abiotic processes. Note, however, that the latter phrase adds environmental stresses ("keystone processes") such as fire, wave-induced damage, substratum movement, and other kinds of disturbances to the biological forces stressed by the other concepts. A contrasting view, that biotic and abiotic forces are qualitatively distinct, suggests the term "critical processes" to describe abiotic effects that can structure communities.

C. Recommended Terminology

Although the various generalizations of the keystone concept and the lumping of several related concepts into a broad keystone species definition were well intentioned, we agree with Mills *et al.* (1993) and Power *et al.* (1996) that such usage sharply reduces the utility of the keystone species concept. On the basis of current usage and the documented need and desire for clarity in standardized terminology, we propose the adoption of the following terminology. Species in communities can be defined as:

- *Keystone species:* Consumers having a disproportionately large effect on communities and ecosystems. By this definition, keystone species can include predators, parasites, pathogens, herbivores, pollinators, and mutualists of higher trophic status but not plants, sessile animals or "resources" (e.g., salmon carcasses, salt licks, deep pools). To date we know of no convincing examples of communities with more than a single keystone species. We therefore suggest the hypothesis that most *communities* will have at most a single keystone species. At the *ecosystem* level, which might include several distinct communities, there may be several keystone species. Evaluation of these predictions awaits future research.
- *Strong interactors (critical species):* Species having a large effect on the species (one or more) with which they interact. Communities and ecosystems may have many strong interactors, and such species may occur at all trophic levels. Strong interactors would lie toward the upper side of the abundance-impact diagram (Fig. 7), and therefore include both keystone species and dominants.
- *Weak interactors:* Species having little effect on other species, at least under average conditions. Under some circumstances, weak interactors may occupy important roles in ecological communities as a result of changes that lead to temporary increases in abundance, size, or biomass (Berlow, 1999). Weakly interacting species would all lie toward the lower portion of Fig. 7.
- *Dominant species:* As noted above, dominant species are those strongly interacting species that owe their influence to their high abundance (Fig. 7). Such organisms are the species that comprise a large proportion of the biomass in a community, and thus are the dominant components of community structure. Trees in forests, mussels in rocky intertidal habitats, grasses in grasslands, and kelps or corals in nearshore subtidal habitats are all dominants.
- *Key-industry species:* As defined earlier, key-industry species are prey that support a large group of consumers. Following Elton's (1927) usage, key-industry species are therefore animals of intermediate trophic status.

Groups of species in communities or ecosystems can include:

- *Interaction webs:* Interaction webs (= functional webs) are the subset of species that through their interactions and responses to abiotic factors make up the dynamic core of food webs or communities. Interaction webs include keystone species, dominants, and other strong interactors.

III. IDENTIFICATION OF KEYSTONE SPECIES

Lack of experimentation and other rigorous approaches to identify keystones continues to be a pervasive problem. In many cases, species are named keystones based on superficial evidence such as natural history observations. A lack of rigor in identification could result in mislabeling a species as a keystone and inferring that it is the primary determinant of the structure of its community when it is in fact not. As noted by Mills *et al.* (1993), this can yield a serious loss of credibility with respect to policy and management decisions.

Despite limitations of spatial and temporal scales and other shortcomings, experimentation supplemented by comparison remains the most powerful tool available for revealing the dynamics of communities and ecosystems (Paine, 1994). In many cases, however, practical, legal, and ethical concerns preclude manipulations of suspected keystones (e.g., sea otters, killer whales, or prairie dogs). It is clear, therefore, that identification of keystone species must necessarily use a variety of approaches. In addition to the comparative-experimental approach, alternative approaches include path analysis, sensitivity analysis, the study of natural or accidental invasions, the study of the consequences of overexploitation, and exhaustive and detailed comparison coupled with natural history and ideally small-scale ancillary experimentation (Power *et al.*, 1996). Inferences based solely on descriptive natural history knowledge (e.g., diet composition and frequencies, behavior, abundance) are likely to be misleading. For example, under natural conditions a keystone predator may rarely consume the competitive dominant in a system (because it has sharply reduced the prey's availability) and thus be overlooked as a possible regulator of the dominant or the community. In the sea star system studied by Paine (1966), for example, barnacles, not mussels, were the most frequently consumed prey. Thus, although natural history knowledge is fundamental to the understanding of the dynamics of any ecological system, much additional evidence is necessary before the ecological role of a species can be determined.

A. Experimental Approaches

Keystone species are often identified by removal or exclusion experiments; that is, the presumed keystone is deleted from a portion of the habitat, and the effect of the removal on the community is compared to a separate, control portion of the habitat. One problem with this approach is that if a suspected keystone species is removed but there is no detectable response by the community, it is not possible to conclude that the community lacks a keystone species. In such a case, three alternative interpretations are possible. (1) Predation is weak overall such that no predator deletion will produce an effect. (2) Predation is strong and there is a keystone species in the community, but it was not the species removed. (3) Predation is strong but diffuse, and multiple predators would need to be removed to produce an effect.

To tease apart these alternatives, an appropriate experimental protocol in testing for keystone predation (Menge *et al.*, 1994; Navarrete and Menge, 1996) should include:

• Treatment	• Explanation	• Tests
• 1. +Predators	• Intact community, all predators present	• Control or reference ("natural" community state)
• 2. −Predators	• Quantifies total predation effect, all predators removed or excluded	• Strong versus weak predation
• 3. −Single Predator	• Quantifies single-species effects, deletion of each predator species singly, leaving the others present	• Keystone versus diffuse predation

For example, if consumer species are removed both collectively and singly, but predation is weak, the total effects of predation and the effects of each species with respect to their impact on community structure will be small (Fig. 8a). With strong predation regimes, keystone predation would be indicated if total removals and removal of just one of the predator species were similarly large (Fig. 8d). Finally, diffuse predation would be indicated if removal of each predator species

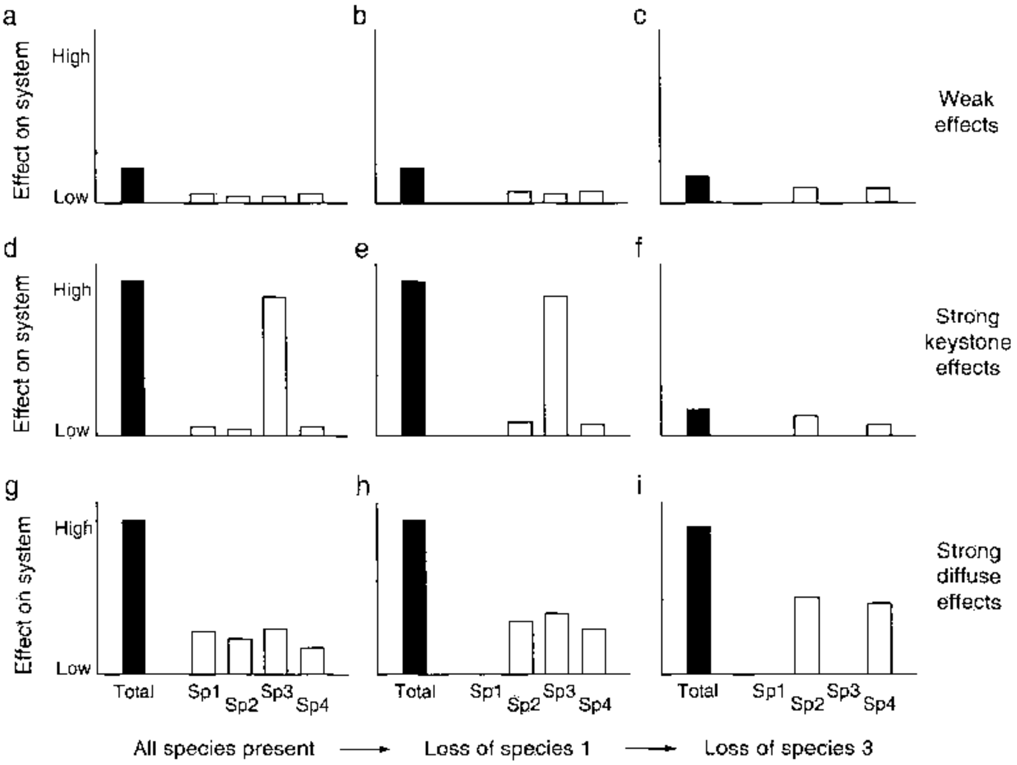


FIGURE 8 Total effects (solid bars) of a consumer group and effects of single species (open bars) under three regimes of predation effect, with possible responses to sequential loss of single species. Panels a, d, and g represent the intact assemblage; b, e, and h represent the changes in the absence of species 1; and c, f, and i represent the changes if both species 1 and 3 are removed or lost. Panels a, b, and c show change with weak predation; d, e, and f show changes with strong keystone predation; and g, h, and i show changes with strong diffuse predation. In the absence of the keystone species, the remaining species have a small impact on the community (f). In contrast, with diffuse predation, the remaining species compensate strongly for the loss of a consumer so that predation remains strong. (From Allison *et al.*, 1996. Copyright John Wiley & Sons. Reproduced with permission.)

singly led to significant but substantially smaller effects than removal of all predator species (Fig. 8g).

This protocol also has important implications for understanding how communities or ecosystems might respond to losses of trophically high species (Allison *et al.*, 1996). In a system where predation is weak, of course, loss of species should have little effect (Fig. 8a–c). With strong diffuse predation, where each consumer has an impact on community structure, species loss may lead to little change in the system as a result of compensation by the remaining consumers, at least until most consumers have been lost (Figs. 8g–i and 9). In both cases, the community response to species deletion is relatively predictable (Allison *et al.*, 1996). With strong keystone predation, however, the consequence of species loss is relatively uncertain (Figs. 8d–f and 9). By definition, compensation by the weakly interacting predators, while possible, is not likely to fully account for the loss of the keystone species, and if the identity of this species is not known a priori, the system response is highly uncertain (see Fig. 9).

This protocol has rarely been used, but two examples illustrate its efficacy. Menge *et al.* (1986) evaluated the separate and combined impacts of most combinations of four groups of consumers in rocky intertidal communities in Panama. The extremely high diversity of this tropical community imposed an immediate compromise in their design: simultaneous single *species* removals were essentially impossible because >40 consumer species were relatively abundant. The compromise was to remove major consumer *groups* (i.e., omnivorous crabs, omnivorous fishes, predatory whelks, and grazing molluscs), each consisting of several common species. Their partial factorial design (difficulties in separating crab and fish effects meant only 12 of a possible

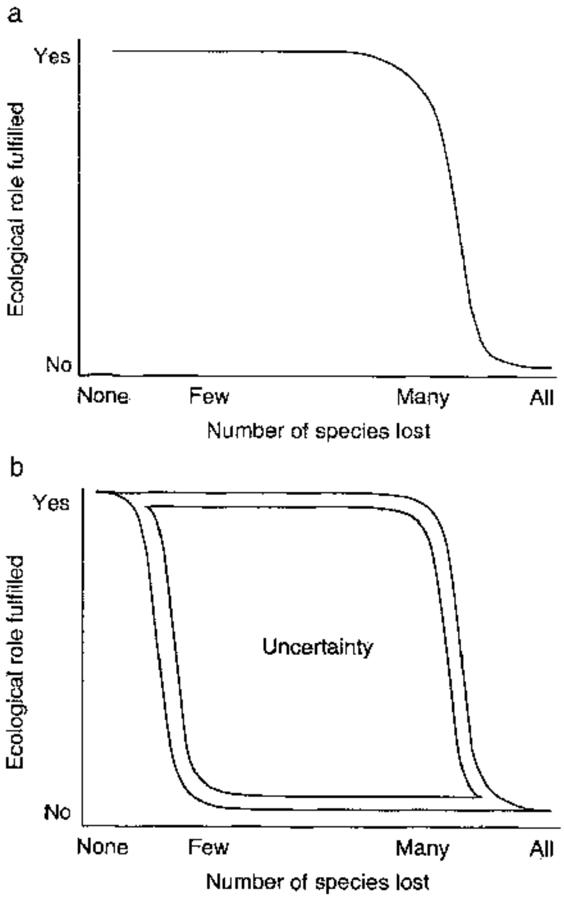


FIGURE 9 Relationship between the ability of a group of consumers to maintain a system at some persistent state ("ecological role fulfilled") and the regime of predation. (a) Diffuse predation: in moving from left to right on the abscissa, the remaining species compensate for the loss of other species, maintaining similar structural patterns in the community or ecosystem until most are lost. (b) Keystone predation: as species are lost, dramatic changes in structure can occur with each removal, but unless the identity of the keystone species is known, prediction of when such changes occur is uncertain. (From Allison *et al.*, 1996. Copyright John Wiley & Sons. Reproduced with permission.)

16 treatments could be conducted) included treatments assessing the effects of removing consumers in single groups and in total. After three years, none of the single-group removals produced an effect similar in magnitude to excluding all the consumers (Fig. 10). Removals of two groups and three groups produced changes in prey that were intermediate between single-group and total exclusion treatments. Thus, total predation pressure in this system was high but the effect of individual groups was small. The lack of prey responses from single-group removals compared to the increasingly strong responses from removals of two, three, and all four groups suggests three things: that, up to a point, groups can compensate for reductions in other groups; that there was no keystone species in this system; and that predation was diffuse.

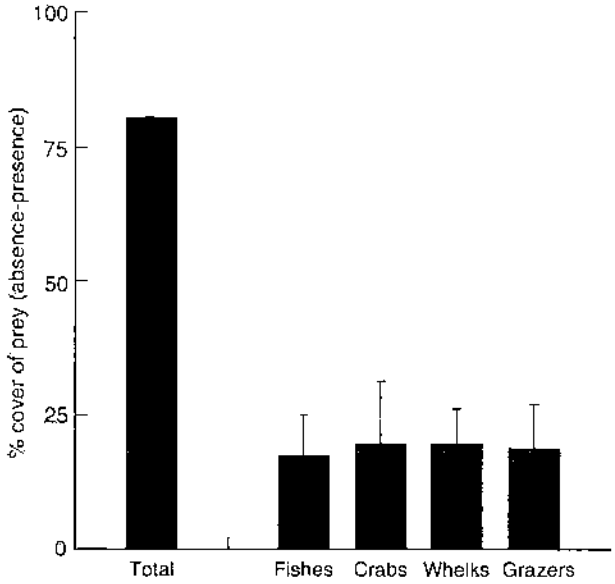


FIGURE 10 Effect of consumers, collectively and by each of four single groups (fishes, crabs, whelks, molluscan grazers), on total prey abundance in the rocky intertidal region at Taboguilla Island, Panama. Prey included sessile invertebrates (mostly barnacles and bivalves), colonial invertebrates (hydrozoans), and algae. The estimate of the effect of consumers was based on the percent cover of prey in the absence and presence of all consumers (total) and each group separately. Single group effects were estimated by determining the difference between treatments that differed in whether or not the particular group was included in the treatment. For example, fish effects were estimated by treatments +F+C+W+G versus −F+C+W+G, +F+C−W+G versus −F+C−W+G, etc. Error bars are 1 SE. (From Menge *et al.*, 1986.)

In another study, Navarrete and Menge (1996) tested the relative intensities of predation on the small mussel *Mytilus trossulus* by the original keystone species, the sea star *Pisaster ochraceus*, and coexisting predatory whelks, *Nucella emarginata* and *N. canaliculata*, on the Oregon coast. In a fully factorial design they quantified rates of predation on mussels in all combinations of presence or absence of each. Their results indicated that, as predicted, *Pisaster* had the strongest effect, by far, on mussel survival (Fig. 11). Although whelk effects were almost undetectable in the presence of sea stars, they were relatively strong in the absence of the sea stars. Thus, *Pisaster* was the keystone predator in these experiments. Whelks were weak interactors in the presence of sea stars, but, via compensatory increases in density, had moderately strong effects on mussel survival when sea stars were absent.

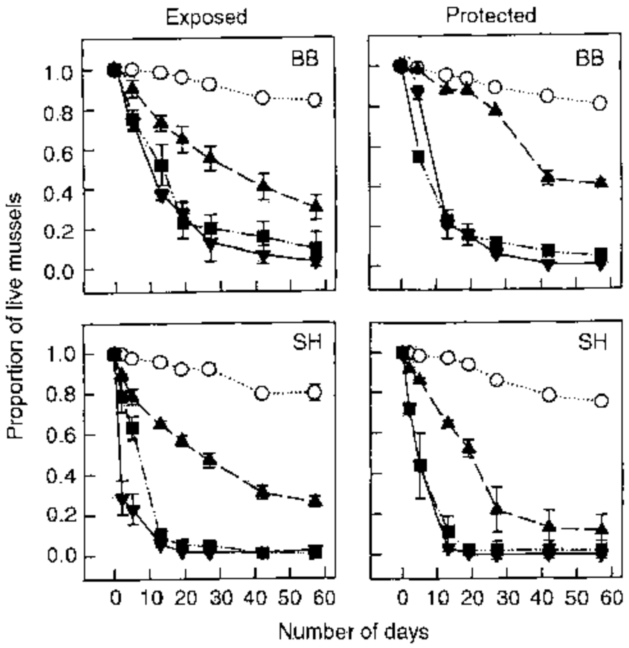


FIGURE 11 Results of a mussel transplant experiment. Data are the mean proportions of mussels (±1 SE) surviving in each of four treatments at wave-exposed and wave-protected sites at two study sites, Boiler Bay (BB) and Strawberry Hill (SH). Treatments were all combinations of presence and absence of whelks (*Nucella* spp.) and sea stars (*Pisaster*). ○, −*Pisaster* −*Nucella*; ▲, −*Pisaster* +*Nucella*; ■, +*Pisaster* −*Nucella*; ▼, +*Pisaster* +*Nucella*. The experiment was begun 3 July 1993 and completed 30 August 1993. (From Navarrete and Menge, 1996.)

B. Nonexperimental Approaches

To be convincing, comparative evidence needs to be extensive in space and time and needs to incorporate a wide range of approaches. One of the most convincing examples of the determination of keystone species on the basis of description and comparison is the sea otter example described earlier. Another convincing example is that of beavers as keystone species (Naiman *et al.*, 1986). By building dams across streams and through feeding activities, beavers can have profound effects on the structure and dynamics of aquatic ecosystems. Beavers alter hydrology, channel geomorphology, productivity, and biogeochemical cycling at a magnitude far exceeding their proportional biomass in these systems.

Bison are another good example of keystone species designation based on largely nonexperimental evidence (Knapp *et al.*, 1999). Extensive and detailed studies at the levels of individual leaves, individual plants, plant populations, and landscapes showed that selective grazing at the level of species and patch greatly altered patterns of community structure. Bison preferred grasses in burned areas, and revisited these sites repeatedly, establishing a distinct, high-diversity mosaic pattern of vegetation. Their grazing on grasses and avoidance of forbs resulted in high levels of species diversity, high spatial heterogeneity, and high nitrogen availability in tallgrass prairie. Moreover, bison grazing interacted with fire, another natural element of tallgrass prairies, in ways that enhanced the patchiness, diversity, and nitrogen availability, and generated intermediate levels of net primary productivity. As mentioned earlier, estimates of the CI of bison indicate that these large grazers clearly have a large effect, one disproportional to their biomass, on plant community structure and ecosystem functioning.

Space does not permit further detailed citation, but other convincing examples of keystone species are available from marine, freshwater, and terrestrial habitats, including those cited in Table 1 in Power *et al.* (1996). Many more examples than these are available in the literature that are based on limited evidence and thus remain conjectural. It seems clear, however, that the phenomenon of keystone species is real and widespread, and that a variety of approaches are available to document their existence and importance. What remains unclear from the foregoing is (1) whether or not there are keystone traits that allow a priori identification of keystone species, and (2) whether or not there are predictable conditions under which keystone species will evolve and persist. We consider these issues next.

C. Identification of Keystone Species a Priori?

Field experimentation is not always feasible. Moreover, even where and when feasible, such studies are often time-consuming and expensive. It seems obvious that we cannot hope to study each and every system on Earth to determine the nature of interactions among species and how these influence community dynamics. For these reasons an ultimate goal in ecology is to gain the ability to predict the roles of species and the processes underlying the dynamics of communities and ecosystems. Ideally, predictions would be based on observation and measurement of organismal, population, and community traits and patterns. Evaluation of the reliability of such predictive traits would depend on cycles of study that identified important traits in particular systems and tested their predictive capacity in novel systems.

Traits originally suggested to characterize keystone species were differential impacts on prey species, high consumption rates relative to prey production, and a

TABLE I

Summary of Survey of Potential Keystone Species Traits in Consumer-Dominated Aquatic Communities

	Number of interaction webs					
	Keystone predation			Diffuse predation		
Trait	Yes	No	?	Yes	No	?
1. Differential predation	7	0	4	3	0	3
2. Predation focused on competitive dominant	9	0	2	4	0	2
3. High predator/prey size ratio	10	1	0	6	4	0
4 High functional/numerical response	8	2	1	6	4	0
5. Indeterminate growth of predator(s)	8	3	0	6	4	0
6. High predator mobility relative to prey	10	1	0	6	4	0
7. Spatial refuge for prey	6	4	1	3	2	1
8. Size refuge for prey	1	8	2	2	3	1
9. Space is limiting resource for prey	7	4	0	5	1	0
10. Low community diversity	3	5	3	2	2	2
11 Differential rates of prey recovery	6	1	4	4	0	2
12. High rate of prey production	5	0	6	0	2	4

Question mark means the conclusion is uncertain or unknown. For traits 3–6 the numbers under diffuse predation total >6 because several of the systems had several strongly interacting predators, each of which had a yes or no or ? conclusion. After Table 10 in Menge *et al.* 1994.

focus of the keystone's impact on dominant competitors in the system. A recent survey of well-studied keystone predator examples in marine and freshwater habitats, however, suggested that these, and several other postulated "keystone-identifying" traits, were not predictably associated with keystone species (Menge *et al.*, 1994; Table I). In all, 17 studies of strong predation were considered. Eleven were dominated by keystone predation and six were dominated by diffuse predation. As indicated in Table I, differential predation and dominance of the community by a single prey species in the absence of the predator were characteristic of both keystone and diffuse predation systems, and thus do not serve as a distinguishing trait of keystone predation. Although data were limited, prey production rate was the only trait that consistently distinguished keystone (high prey production) and diffuse (low prey production) predation systems. No other potential trait was consistently associated with either type of system. Clearly further research is needed, but based on present knowledge, a priori identification of keystone species using traits alone does not seem possible.

D. Context Dependency

The role of a species as keystone can be context dependent (Menge *et al.*, 1994; Power *et al.*, 1996), that is, a species that serves a keystone role under some set of biotic and/or abiotic conditions may not under other conditions. For example, at exposed headlands along the Oregon coast, the sea star *Pisaster ochraceus* was found to occupy a similar role to that documented for this species in comparable habitats on the Washington coast (Menge *et al.*, 1994; Paine, 1966). On more wave-sheltered rocky shores, however, *Pisaster's* role was weak to absent, despite the presence of populations of this sea star at all locations studied (Menge *et al.*, 1994). The basis for this change was that environmental conditions changed with diminished wave impact. At one sheltered site, prey production rates (a combination of recruitment rates and prey growth) were so low that the sea star population consisted of a few scattered individuals. This and other studies suggest that sea star abundance is highest where food concentrations are high. At the other sheltered site, periodic and unpredictable burial by sand, not sea star predation, eliminated the competitively dominant mussels (*Mytilus californianus*) from the lower shore (Menge *et al.*, 1994). Keystone predation in this system, then, depended on the spatial and environmental context in which the community occurred.

Keystone effects may also be altered by temporal environmental changes. Recent studies on the Oregon coast showed that *Pisaster* foraging activity is strongly

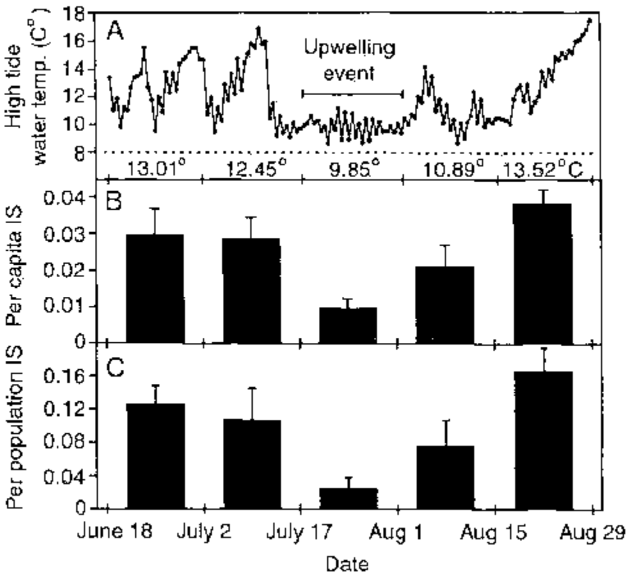


FIGURE 12 Seawater temperature and sea star–mussel interaction strength (IS) during consecutive 14-day periods (abscissa). Data are means (+1 SE) of values recorded at three sites. (A) High tide water temperatures (the mean from two hours before to two hours after each high tide). Temperatures above the abscissa are the overall means for the 27 high tides of each period. (B) Mean per capita IS of sea stars on transplanted mussels during each 14-day period. (C) Mean per-population IS of sea stars during each period. Both per-capita and per-population IS were lower during upwelling than during the other four periods. (From Sanford, 1999.)

suppressed by cold water temperatures such as those occurring during periodic upwelling (Sanford, 1999; Fig. 12). Although further study is needed to determine the broader significance of this result, sea temperature variation in time or space could also modify the community role of this species. In particular, alteration of upwelling patterns by global climate change could lead to community- and ecosystem-level modifications or shifts in community pattern over a broad geographic region. Thus, investigations conducted over a wide range of environmental conditions indicated that whether or not a species can fill a keystone role is dependent on physical and biological conditions (i.e., context).

In a second example of context-dependent keystone predation, juvenile steelhead were demonstrated to control summer food web structure in California rivers in years experiencing winter flooding. Steelhead were the top predator in a four-level food chain, and when present they controlled invertebrate predators, freeing herbivorous insects from consumer control and allowing them to control the abundance of algae (Power, 1990, cited in Power *et al.*, 1996). When steelhead were excluded, invertebrate predators increased and controlled herbivorous insects, resulting in a high biomass of algae. Winter floods swept these communities away, leading to annual cycles of community redevelopment, with shifts in algal biomass reflecting the development of successive trophic levels through the summer into autumn.

In years without winter flooding, predation-vulnerable herbivorous midge larvae (*Pseudochironomis*) were replaced by predator-invulnerable herbivores (e.g., aquatic moth larvae *Petrophila*, caddisflies *Tinodes* and *Glossosoma*) (Power, 1992, 1995, cited in Power *et al.*, 1996). The invulnerable herbivores sharply suppressed algal abundance and, due to their effective defenses against predators, effectively shortened the food chain from four to two levels. Thus, fishes (juvenile steelhead) were keystone predators in the context of years with winter flooding and were not in the context of years without winter flooding.

A variety of similar examples are available from all major habitats (see Table 2 in Power *et al.*, 1996), suggesting that context dependency is common and perhaps near-universal in ecological communities. Such variation may be a useful tool in investigating the conditions under which keystone species will evolve and persist. Understanding the conditions under which a species plays a keystone role is critical if ecologists are to build a predictive framework to evaluate the effects of global changes in biodiversity.

E. Interaction Strength

As noted earlier, keystone species are distinguished by both the strength of their interactions with other species and the large indirect consequences of these effects. Since determination of whether or not a species is a keystone depends on rigorous application of methods that can quantify these effects, it is important to briefly consider the meaning of the concept of interaction strength and the methods used to estimate it.

The concept of interaction strength has different meanings depending on the context. Theoretical models, such as the Lotka–Volterra models of interspecific competition, include a coefficient, α_{ij}, which "measures how strong the interactions are" (MacArthur, 1972). This coefficient reflects the per-individual or "per-capita" effect of one species *j* on another, species *i*. Similar coefficients are included in predator–prey models as well. MacArthur (1972) also suggested a second meaning of interaction strength: interactions can be considered as strong if their "removal would produce a dramatic effect." In this case, interaction strength refers to population-level, not per-capita, impacts of one species on others.

Despite its long history in ecology, the concept of interaction strength has only recently been explored by experimental field ecologists. Paine (1992), for example, quantified per-capita interaction strengths in field enclosure/exclosure experiments involving herbivores (sea urchins, limpets, and chitons) grazing on kelp sporelings. Specific impacts of individual herbivore species were determined by comparing the densities of kelp sporelings that occurred in the complete absence of all herbivores (a measure of total herbivore pressure) to those occurring in the presence of single herbivore species (a measure of a single species' effect). Per-capita effects were estimated by dividing by the number of herbivores used in each single-species treatment.

This analysis showed that per-capita species interaction strengths varied among the species of this herbivore assemblage. Three interactors were described as "strong." Sea urchins and a large chiton had large negative effects on algae. A smaller chiton had relatively large positive effects on kelp sporeling densities in one experiment and weak effects in another, showing that interaction strength can vary within species. The other four species had no distinguishable per-capita effects and were therefore weak interactors. Similar efforts have been made with preying mantis interactions with its insect prey, with predation effects on muddy intertidal community structure, and with bird predation on rocky intertidal invertebrates (see Berlow *et al.*, 1999, and citations therein).

One important outcome of these studies, consistent with expectation, was that these communities or assemblages consisted of a few strong and many weak interactors (Paine, 1992). More broadly, this approach is a potentially powerful tool in distinguishing systems with weak predation from those with strong predation. Further, in those systems with strong predation, this method should separate keystone predation regimes from "diffuse" predation regimes.

Quantification of interaction strength is sensitive to the particular metric used in its estimation and to the particular theoretical concept of interaction strength (Berlow *et al.*, 1999). Estimates of per-capita effects can vary with density of the prey and predator, with the proximity of the system to equilibrium, and with the particular index used. For example, both the CI index given earlier and the index used by Paine (1992) yield measures that are asymmetric about zero, inflating either negative (CI) or positive (Paine's index; PI) values in one direction but bound at a maximum of 1 (or -1) in the other direction (Berlow *et al.*, 1999). Another index, the "dynamic index" (DI; an index that is based on the rate of change immediately after a perturbation), provides symmetrical estimates of interaction strength bounded at -1 and $+1$. The CI and PI indices provide consistent estimates when abundances are at or near equilibrium but DI does not; when abundance is changing, DI provides consistent estimates but CI and PI do not. It is thus important that investigators pay close attention to the specific experimental protocol, the underlying model for the interactions studied, and the type of index used.

Using these approaches, keystone species will be those with large per-capita effects. Identifying such interactors in this way might shed light on the traits, both individual and community, that generate species with high per-capita effects. As implied by the alternative meanings for interaction strength offered by MacArthur (1972), however, species impacts will also vary with density. Regardless of how large its per-capita effect, a single individual will never have the same impact as that of many individuals (Berlow *et al.*, 1999; Navarrete and Menge, 1996). Thus, fuller understanding of interaction web dynamics, and whether a species is a keystone, a strong interactor, a weak interactor, or a dominant, will depend on knowledge of both per-capita interaction strength and population interaction strength or "species impacts." Note also that only a keystone species will have high values of both per-capita and per-population indices of interaction strength, at least at the community or ecosystem level.

IV. CONSERVATION IMPLICATIONS

Biologists are being increasingly asked to inform management decisions concerning the preservation of biological diversity and ecosystem "integrity." Because keystone species can be critical to the maintenance of species diversity and ecosystem functioning, some have argued that focusing conservation on them should be a priority. A current topic of debate is how to best design reserves, both marine and terrestrial, that will protect species diversity.

Many species have been presumed to be keystones based on observation and conjecture. In many cases no experimental removals were done to determine the actual effect of the species presumed to be a keystone on the rest of the community. As outlined earlier, identification of keystones should involve manipulations to quantify the effect of the suspected keystone on the community. However, from a management standpoint this is not always feasible. It sometimes takes years for field manipulations to yield results, while conservation decisions often need to be made in a much shorter period of time. Also, as noted earlier, in addition to potential time constraints, manipulations of some suspected keystones may not be practically feasible or ethical. Removing killer whales, sea otters, lions, polar bears, or other large, charismatic, and often endangered species is not a realistic option. The logistic difficulties in designing controlled exclosure or removal experiments for animals of large size and/or with large ranges are daunting.

Despite the difficulties, and accepting the view that conservation should focus on ecosystems rather than species, the keystone species concept offers several important insights that are relevant to management (Power *et al.*, 1996):

- Seemingly scarce and unimportant species may have unexpectedly large, dramatic effects on communities and ecosystems;
- Conserving species may depend strongly on other species in the community with which the target species has little seeming association, whether as prey, competitor, mutualist, or predator; and
- Loss of a species, particularly one high in the food chain, may have surprising and extensive consequences for the remainder of the community or ecosystem.

All of these points suggest that great caution is necessary before decisions are made that may result in the loss of a native species or the introduction of exotic species. We recognize, and agree with, the need to make management and policy recommendations on the basis of present knowledge. We also argue that, although much has been learned and the pace of new knowledge is increasing, the current state of knowledge is still insufficient to allow making such recommendations with a high level of confidence. In particular, we need additional study on those putative keystones for which evidence is scant, and continued study on ways to detect keystone species and systems with keystone species. Other urgent issues are to determine the commonness of keystone species and keystone-dominated systems, the patterns of interaction strengths in representative communities, and the environmental and organismal conditions and traits that foster keystone species.

We conclude that the concept of keystone species is a powerful one, having broad importance and application to ecological theory, ecosystem dynamics, and conservation. Many problems remain to be answered, but we have made great progress and are in the midst of a period of dramatic advances on these issues.

Acknowledgments

We thank Kealoha Freidenburg, Heather Leslie, Bob Paine, Mary Power, Eric Sanford, David Skelly, and Michael Webster for constructive criticism. This article was written while the authors were supported by grants from the Andrew Mellon Foundation (B.A.M.), the Wayne and Gladys Valley Foundation (B.A.M.), the David and Lucile Packard Foundation (B.A.M.), and an NSF Graduate Fellowship (T.L.F.). Contribution number 9 from PISCO, the Partnership for Interdisciplinary Studies of Coastal Oceans: A Long-Term Ecological Consortium, funded by the David and Lucile Packard Foundation.

See Also the Following Articles

COMPETITION, INTERSPECIFIC • CONSERVATION BIOLOGY, DISCIPLINE OF • ECOLOGY, CONCEPTS AND THEORIES IN • ECOSYSTEM FUNCTION, PRINCIPLES OF • FOOD WEBS • PREDATORS, ECOLOGICAL ROLE OF • SPECIES COEXISTENCE

Bibliography

Allison, G. W., B. A. Menge, J. Lubchenco, and S. A. Navarrete. (1996). Predictability and uncertainty in community regulation: Consequences of reduced consumer diversity in coastal rocky ecosystems. In *Functional Roles of Biodiversity: A Global Perspective* (H. A. Mooney, J. H. Cushman, E. Medina, O. Sala, and E.-D. Schulze, eds.), pp. 371–392. John Wiley & Sons, Chichester, United Kingdom.

Berlow, E. L. (1999). Strong effects of weak interactors in ecological communities. *Nature* 398, 330–334.

Berlow. E. L., S. A. Navarrete, M. E. Power, B. A. Menge, and C. Briggs. (1999). Quantifying variation in the strengths of species interactions. *Ecology* 80, 2206–2224.

Brown, J. H. (1998). The desert granivory experiments at Portal. In *Experimental Ecology: Issues and Perspectives* (W. J. Resetarits, Jr., and J. Bernardo. eds.), pp. 71–95. Oxford University Press, Oxford, United Kingdom.

Dayton, P. K. (1972). Toward an understanding of community resilience and the potential effects of enrichments to the benthos at McMurdo Sound, Antarctica. In *Proceedings of the Colloquium on Conservation Problems in Antarctica* (B. C. Parker, ed.), pp. 81–96. Allen Press, Lawrence, Kansas.

Elton, C. S. (1927). *Animal Ecology*. Sidgwick & Jackson, London.

Estes, J. A., N. S. Smith, and J. F. Palmisano (1978). Sea otter predation and community organization in the western Aleutian Islands, Alaska. *Ecology* 59, 822–833.

Knapp, A. K., J. M. Blair, J. M. Briggs, S. L. Collins, D. C. Hartnett, L. C. Johnson, and E. G. Towne. (1999). The keystone role of bison in North American tallgrass prairie. *BioScience* 49, 39–50.

MacArthur, R. H. (1972). Strong, or weak, interactions? *Trans. Conn. Acad. Arts Sci.* 44, 177–188.

Menge, B. A., E. L. Berlow, C. A. Blanchette, S. A. Navarrete, and S. B. Yamada. (1994). The keystone species concept: Variation in interaction strength in a rocky intertidal habitat. *Ecol. Monographs* 64, 249–286.

Menge, B. A., J. Lubchenco, L. R. Ashkenas, and F. Ramsey. (1986). Experimental separation of effects of consumers on sessile prey in the low zone of a rocky shore in the Bay of Panama: Direct and indirect consequences of food web complexity. *J. Exp. Marine Biol. Ecol.* **100**, 225–269.

Mills, L. S., M. E. Soulé, and D. F. Doak. (1993). The keystone-species concept in ecology and conservation. *BioScience* **43**, 219–224.

Naiman, R. J., J. M. Melillo, and J. E. Hobbie. (1986). Ecosystem alteration of boreal forest streams by beaver (*Castor canadensis*). *Ecology* **67**, 1254–1269.

Navarrete, S. A., and B. A. Menge. (1996). Keystone predation and interaction strength: Interactive effects of predators on their main prey. *Ecol. Monographs* **66**, 409–429.

Paine, R. T. (1966). Food web complexity and species diversity. *Amer. Naturalist* **100**, 65–75.

Paine, R. T. (1992). Food-web analysis through field measurement of per capita interaction strength. *Nature* **355**, 73–75.

Paine, R. T. (1994). Marine rocky shores and community ecology: An experimentalist's perspective. In *Excellence in Ecology* (O. Kinne, ed.). Ecology Institute, Oldendorf/Luhe, Germany.

Power, M. E., D. Tilman, J. A. Estes, B. A. Menge, W. J. Bond, L. S. Mills, G. Daily, J. C. Castilla, J. Lubchenco, and R. T. Paine. (1996). Challenges in the quest for keystones. *BioScience* **46**, 609–620.

Sanford, E. (1999). Regulation of keystone predation by small changes in ocean temperature. *Science* **283**, 2095–2097.

LAKE AND POND ECOSYSTEMS

Christian Lévêque
Institut de Recherche pour le Développement (IRD) Paris

GLOSSARY

ancient lakes Lakes with a persistence of more than 100,000 years are called long-lived or ancient lakes.
eutrophication The process of enrichment of a water body due to an increase in nutrient loading.
species flocks An aggregate of closely related species that share a common ancestor and are endemic to a geographically circumscribed area.

COMPARED TO THE SEA, freshwaters are deficient in major taxa and there are no uniquely freshwater metazoan phyla (May, 1994). In river lakes, aquatic biota is similar to the biota of the river basin. The great majority of existing isolated lakes (around 10,000 exceeding 1 km^2 in extent) are geologically very young and their flora and fauna are usually depauperate compared to ancient lakes that exhibit a rich endemic fauna for several major groups of animals. Species flocks for fish and invertebrates are known from ancient lakes and represent a unique biological heritage to be preserved. Major threats to the lakes biota include habitat alteration, fisheries practice, pollution, and the introduction of exotic species.

I. BROAD CHARACTERISTICS OF THE BIODIVERSITY IN LAKES AND PONDS

Freshwater habitats are widely considered to be transient in time and space in comparison with both terrestrial and marine habitats. This is true for many lakes and ponds. However, depending on the origin of lakes there are great differences in the nature and the diversity of their biota. Three broad categories of lakes may be recognized:

1. *Lakes and ponds that are permanently or frequently connected to large river systems.* This category includes river-lakes (i.e., Lake of Geneva) or lakes that are part of a large floodplain system such as "varzea" lakes. In these lakes, exchanges of flora and fauna occur with the main river system so that their biota is usually greatly similar to the biota of the river system itself with the exception of a few species adapted to still waters. Many endoreic lakes (Lake Chad, Aral Sea) also belong to this group.
2. *Isolated lakes with a limited drainage system.* The biota of the lakes in this category has evolved in isolation from others for a more or less long period of time leading to speciation and endemicity when the period

Encyclopedia of Biodiversity, Volume 3

is long enough. The associated ice ages at higher latitudes and altitudes were the phenomena that created most of the lakes in existence today. Therefore the great majority of existing lakes (around 10,000 exceeding 1 km^2 in extent) are geologically very young and occupy basins formed by ice masses or glacial erosion after the retreat of continental ice sheets some 10,000 years ago. All such lakes are expected to fill slowly with sediment and to disappear in the future, along with any isolated biota. Compared to ancient lakes, they acquired their fauna and flora via the rivers that supply them with water as a result of runoff in their basin and from aerial transport by wind or animals. Only a few existing lakes are known to be much older, and most of them occupy basins formed by large-scale subsidence. They may date back at most 20 million (Lake Tanganyika) or 30 million (Lake Baikal) years. These so-called ancient lakes are of particular interest for biodiversity because they exhibit a rich endemic fauna for several major groups of animals. There is also good evidence that some extinct lakes were also very large and long-lived under different climatic and tectonic conditions.

3. *Temporary lakes and ponds whose water budget is controlled by the climate regime.* The fauna and flora in these lakes exhibit special biological adaptations to seasonal drying.

A. Origin and Peculiarities of Freshwater Biota

It is thought that the early evolution of all the major animal phyla took place in the sea. Most phyla are predominantly marine and benthic: 32 phyla are found in the sea with 11 exclusively marine, whereas 14 are represented in freshwater and only 12 are found on land (May, 1994). Compared to the sea, freshwaters are deficient in major taxa and there are no uniquely freshwater metazoan phyla. The osmotic challenges of life in freshwaters probably discouraged invasion of the habitat by many marine invertebrates. It explains probably the tendency in freshwater invertebrates for larger but fewer eggs than in marine relatives: they must eclose with fully developed osmoregulatory capacities to be at a more advanced stage to cope with the highly dilute surrounding.

Another difference between the species richness of marine and freshwater zooplankton derives from the necessity of diapause or other resting mechanisms as a condition for persistent successful radiation in freshwaters (Lehman, 1988). Freshwater invertebrates developed anabiotic devices: special resistant eggs, cysts, and other resting stages that are produced to tide the animal over periods of desiccation, extreme cold, heat, anaerobic situations, lack of food, and other adverse conditions. In addition to withstanding unfavorable conditions, resistant stages have the further function of making overland transport and geographical dissemination possible. Without such a function, colonization of freshwater areas would be slow and difficult in the discontinuum of isolated lakes and ponds.

B. The Latitudinal Gradient

It is usually assumed that species diversity increases from high to low latitudes for most of the major groups of plants and animals and that highest values occur at low latitudes. Indeed, the diversity of marine plankton decreases from low latitudes to high ones, so that tropical and subtropical ocean waters exhibit rich diversity of zooplankton whereas Arctic and Antarctic waters tend to be dominated by copepods and euphausiids. In freshwaters however, the latitudinal trend in species richness is the opposite. Tropical lakes have abbreviated zooplankton faunas compared with temperate locales (Fernando, 1980); they are depauperate in large-bodied species of copepods and Cladocera, and limnetic rotifers are likewise poorly represented.

It could be assumed that the associated ice ages at higher latitudes and altitudes were the phenomena that created most of the lakes in existence today. They are therefore very young compared to ancient lakes, and they acquired their fauna and flora via the rivers that supply them with water via runoff in their basin and via aerial transport by wind or animals.

For fish, the species richness is actually smaller in north temperate lakes of glacial origin than in long lived lakes from tropical areas. At least the endemicity is much lower in temperate lakes than in tropical lakes.

Dumont (1994), in a review of the species richness of the pelagial zooplankton in ancient lakes, provided also evidence that these water bodies have simple pelagial communities. Among 14 pre-Pleistocene lakes across the world, at least one Cyclopoid copepod species is present, often in the genus Cyclops or Mesocyclops, a group of microraptorial species feeding on rotifers, small Crustacea, and immature stage of other copepods.

The number of species regularly found in the pelagic plankton of ancient lakes (pre-Pleistocene) varies from 3 (Lake Tanganyika) to approximately 15 to 20 (up to 5 copepods, 5 cladocerans, 10 rotifers) in Lakes Victoria, Biwa, and Titicaca. In contrast, "young" lakes may have up to 10 species of copepods, 10 of Cladocera, and 10 to 15 species of rotifers occurring together. In the oldest lakes (Baikal, Tanganyika), which also happen to be the deepest, this simplification has gone extreme and the food web reduces to a linear chain.

The question has been raised as to why Cladocera have been almost completely eliminated from some ancient lakes such as Tanganyika, Baikal, and even Malawi (Dumont, 1994). They are able to eat both large items and microplankton and seem all but competitively inferior to other species for food acquisition. Predation had been advocated as a possible cause. In clear-water lakes such as Tanganyika and Baikal, visual predation by fish is more effective than in turbid lakes, and large clumsy swimmers like big Daphnia are likely to be preyed to extinction before the relative small transparent, agile swimmers like the Calanoids. An experimental demonstration of this hypothesis has been the disappearance of all Cladocera from the pelagial of Lake Kivu, within a decade, after the introduction of the zooplanktivore clupeid Limnothrissa miodon, native from Lake Tanganyika (Dumont, 1986).

C. Vertical Distribution in Lakes

Aquatic organisms are not evenly distributed along depth. Water characteristics are relatively uniform in shallow lakes, which are mixed by winds. However, deeper lakes exhibit patterns of vertical gradients for temperature and light. Briefly speaking, the lake is divided by a thermocline into an upper layer, the epiliminion, and a lower layer, the hypolimnion. Life occurs in the oxygenated upper layer while the lower layer is deoxygenated. In most stratified lakes therefore, biota is very depauperated, except for bacteria, below a few 10 meters in depth. There is an exception, Lake Baikal, which is the deepest lake in the world with a maximum depth of 1620 m. The mechanism of mixing of the deep water zone is still not completely understood, but the entire water column is well oxygenated. Life for fish and invertebrates is therefore possible from surface to maximum depth, which is exceptional for freshwater systems. Lake Baikal is therefore unique among inland systems to the study of bathymetric segregation and includes some of the deepest occurring freshwater animals. Among fishes, the family Abyssocottidae contains 20 species distributed throughout the depths of the lake. Species of the genera *Abyssocottus, Cottinella*, and *Neocottus* are adapted to the deep water way of life in that they do not occur above 400 m, the size of the eyes is reduced, and they are physiologically adapted to resisting high pressures (Sideleva, 1994).

The discovery of a deep sea hydrothermal fauna in the 1980s was a surprise for marine biologists. Similarly, hydrothermal vents have been discovered in Lake Baikal, at a depth of 440 m on the sediment floor of Frolikha Bay (Crane *et al.*, 1991), at the foot of an east-west trending fault. Photographs reveal that the center of the vent field is covered by a near-continuous bacterial mat. A white sponge encrusts small cobbles at the periphery of the vent field. Coiled gastropods and whitish translucent amphipods are found among the sponges and on the sediment at the edge of the bacterial mat.

D. Relationship between Species Richness and Area in Lakes

Community ecologists used to compare isolated freshwater systems to biogeographic islands. The relationship of species number to area containing those species is a well-known empirical observation, and a power function is widely used to describe this pattern mathematically: $S = cAZ$, where S is the number of species, A the area, Z the slope of the regression line, and c a constant. It can also be expressed as $\log S = c + Z \log A$.

The effect of lake size on species richness of invertebrates has been demonstrated. For crustacean zooplankton, species richness is also significantly correlated with lake surface area (Dodson, 1992). The species area curve for North American lakes is statistically different from and steeper than the corresponding European curve (slopes, respectively, 0,094 and 0,054). The log species richness is also correlated to log of the average photosynthetic flux per cubic meter and log number of lakes within 20 km of the target lake. For 66 North American lakes, the three variables can be combined in a multiple linear regression model, which explains 75% of the variation in log species richness (Dodson, 1992).

Species richness of aquatic birds also increases with lake size. In Swiss lakes, the species number increase steeply with lake size up to 50 km^2 and species richness depends more closely on lake area in fish eaters and diving ducks than in dabbling ducks (Suter, 1994). Actually lake size explains 70 to 85% of the variation of abundance and species richness in fish eaters and diving ducks but only 64% of species richness in dabbling ducks. In Florida lakes, bird species richness was also positively correlated to lake area and to total water column phosphorus concentration value (WCP) for each lake. The multilinear Log (species richness) = 1.12 + 0.56 Log (Lake area) + 0.12 Log (WCP) and accounts for 77% of the variance in species richness (Hoyer and Canfield, 1994).

E. Morphometry and Species Richness

The species diversity in a lake is a function of the diversity of habitats; the more ecological niches in the lakes, the more species may be expected. The lake's morphometry is basic to its structure: deep, steepsided lakes do not offer as many biotops than shallow, flat lakes. For the latter, most of the lake bottom may be colonized by plants and animals (the benthic flora and fauna), while in deep lakes, only a small part of the lake bottom is colonized. Generally speaking, deep lakes are dominated by planktonic organisms, which are floating or weakly swimming organisms, usually associated with suspended particles. In shallow lakes, benthic organisms are dominant and the heterogeneity of lake bottom, as well as the development of macrophytes, may increase the diversity of benthic species.

II. EVALUATING BIOLOGICAL DIVERSITY

Despite the efforts of taxonomists, a good estimation of the total number of species occurring in freshwater lakes and ponds does not exist. We shall provide here some recent findings about aquatic biodiversity.

A. Diversity of Plankton and Microbial Loop

Three major size classes are usually recognized in pelagic plankton: microplankton (20–200 μm), nanoplankton (2–20 μm), and picoplankton (0.2–2.0 μm). In the late 1970, phototrophic picoplankton was discovered in great abundance in both marine and freshwater ecosystems. However, identifying picoplankton causes considerable taxonomic problems due to the very small sizes of these organisms. We do not know how many bacterial species exists in the world, because bacteria cannot be differentiated under the microscope; we do not even know the right order of magnitude. A new way of classification has been proposed, based on the sequences of ribosomal RNA that led to a phylogenetic classification of bacteria. It is becoming apparent that the genetic diversity among bacteria is much wider than that among the animals and plants.

Most heterotrophic nanoplankters are small (2–5 μm), colorless flagellated protists. They grow at about the same rate as bacteria and are capable of consuming the entire bacterial production. Meanwhile, they regenerate significant amounts of nutrients and serve as prey for micro- and mesoplankton.

The importance of bacteria and protozoa activities in the trophic structure of lacustrine food chain has been largely underestimated in the past. The major role played by microorganisms in controlling energy and nutrient fluxes is now better understood following the discovery of the microbial loop and its role as a source or a sink for carbon and energy flow to higher trophic levels in pelagic systems. We know now that these microorganisms can control major fluxes of energy and nutrients. In some cases, 50% of the photosynthetic production does not pass directly to higher trophic level but is diverted into a microbial loop where nutrients are rapidly remineralized and fed back to the dissolved inorganic pools.

B. Diversity in Freshwater Sediments

About 175,000 species of organisms associated with freshwater sediments have been described, but the true number is much higher than this (Palmer *et al.*, 1997). The number of species in most taxa can scarcely be estimated and global estimate of microbial diversity remains controversial. For example, some specialists estimate that there are hundred of thousands of aquatic nematodes and only a small percent of these have been described. Rotifer species diversity is also poorly known for freshwater sediments, but it is estimated that there are thousands of undescribed species.

Most freshwater sediment species are small and concentrated in the upper sediment layers. Availability of light limits the development of plants and photosynthetic bacteria, which are therefore scarce or absent in most sediments. Moreover, oxygen level may influence species richness and the number of species is low in anoxic waters (see Table I).

C. Diversity in Fish

Presently, 25,000 fish species have been described. Some 10,000 species are found only in freshwaters, a large proportion of which occurs in lakes and ponds. The freshwaters are therefore disproportionately rich in species of fishes on the basis of area when compared to oceans. That could be viewed as the result of the patchy nature of inland waters and the resulting high endemicity of the biota. Fish live in almost every conceivable type of aquatic habitat. They exhibit enormous diversity in size, shape, and biology.

Other vertebrates species occur in freshwaters: a few mammals, several reptiles, and many birds and amphibians. There is no quantitative evaluation of the number of vertebrates whose life cycles include lakes or ponds, but it is far from negligible (see Table II).

TABLE I

Species Richness of the Freshwater Sediment Biota for Many Habitat Types

Taxon	Number of species described	Probable number of species	Range of local species richness
Bacteria	>10,000	Unknown	>1,000
Algae	14,000	>20,000	0–1000
Fungi	600	1,000–10,000	0–300
Protozoa	<10,000	10–20,000	20–800
Plants	1,000	Unknown	0–100
Invertebrates	70,000	>100,000	10–1,000
Aschelminthes	4,000	>10,000	5–500
Annelida	1,000	>1,500	2–50
Mollusca	4,000	5,000	0–50
Acarii	5,000	>7,500	0–100
Crustacea	8,000	>10,000	5–300
Insecta	45,000	>50,000	0–500
Others	1,400	>2,000	0–100

Numbers are rough estimates and derived from many sources. Collected by Palmer *et al.* (1997)

TABLE II

Number of Fish Species Recorded from Several Lakes Connected to Rivers Systems

Lake		Latitude	Area	Number of fish
Chad	Africa	13°N	10–20,000	137
Turkana	Africa	3°N	6,750	51
Chilwa	Africa	15°S	675	31
Ngami	Africa	20°S	150	48
George	Africa	0°	270	30
Huron	North America	44°N	59,600	99
Erie	North America	42°N	25,700	113
Michigan	North America	44°N	58,000	114
Superior	North America	47°N	82,400	67
Great Bear	North America	66°N	31,150	12
Great Slave	North America	61°N	27,200	26
Big Trout	North America	54°N	616	24
Chapala	Central America	20°N	1,080	14
Nicaragua	Central America	11°N	8,200	40
Maggiore	Europe	46°N	676	21
Windermere	Europe	54°N	15	9
Ladoga	Europe	61	18,400	48
Aral sea	Europe	45°N	64,500	17

See also Table III for ancient lakes.

III. BIOLOGICAL DIVERSITY AND ECOSYSTEM FUNCTIONING

Energy and nutrients in an ecosystem are transferred through successive trophic levels. Photosynthesis provides the basic food for herbivorous animals, which are eaten by the carnivores. Therefore, knowledge of the role of individual species and their relationships in aquatic systems is critical to understand the functioning of the system as a whole. Limnologists pointed out several key issues to the study of the relationships between species diversity and lake functioning.

A. Food Webs

Food webs are diagrams depicting which species in a community interact in feeding and describing which kinds of organisms in a community eat which other kind. Food webs are thus caricatures of nature, but they give a picture of the processes at work in ecosystems. Connectivity food webs are describing pathways along which feeding interactions occur. These interactions change at least seasonally and not all interactions are equally strong. The interaction web emphasizes connections that appear to have a large effect on the dynamics of the food web structure and function. Food webs occupy a central position in community ecology. Many important interactions (e.g., competition, predation) cannot be isolated from a food web context because the outcome of these interactions can be modified directly and indirectly by other members of the web.

For a long time food webs served principally as heuristic devices, useful in depicting complex ecosystems as diagrams composed of many interactive parts and enhancing our understanding of pathways of energy and material transfer in aquatic ecosystems. However, the recent surge of interest in food webs seems related to the question of the functional role of biodiversity (discussed later).

Few if any of the aquatic food webs are unimpacted by humans both at a local and a global scale. For example, fisheries food webs are complex, involving multiple trophic levels at several spatial and temporal scales. Fish species offer a wide range of body sizes and feeding habits, and thus have a variety of food web roles and interactions with other species. Exploitation of fishes may result in major changes in food webs. However, the consequences of species removal through fisheries are an almost unexplored field of research in most freshwater systems.

B. The Top-Down Control

In the classical limnological approach, it was usual to consider freshwater ecosystems as operating in a physical-chemical milieu that, largely through nutrient availability, conditions the food chain from primary producers to top predators. In this "bottom-up" control, competition between primary producers for limiting nutrients determines the state of higher trophic levels. A reverse viewpoint, the "top-down control" slowly became prominent. It argues that the effects of top predators cascade down the trophic chain and are responsible for controlling the state of the entire ecosystem. The predators, near or at the top of the trophic pyramid, may be fishes but also may be birds, mammals, and so on, as well as invertebrates. Through grazing, for instance, fish have direct effects on the composition and abundance of phytoplankton, periphyton, and macrophytes, as well as on the dynamics of plankton and benthic communities. Size-selective predation by fish may not only play a major role in the population dynamics of prey species, but also result in shifts in the relative abundance of species.

Predation is now considered to be a major driving force in shaping zooplankton communities. A great number of papers emphasized the size-related alterations in zooplankton communities as a consequence of planktivorous fishes, which select the largest available prey and may rapidly reduce the density of large zooplankters, resulting in a shift of the prey community to small species, predominantly rotifers and small cladocerans. Extinction of large zooplankton has been documented in several habitats, usually following the introduction of new species of planktivorous fish.

Trophic cascades from fishes to water quality in lakes are among the clearest examples of feedbacks from population to ecosystem processes (Carpenter and Kitchell, 1993). A shift in the species composition and size distribution of the fish assemblages alters the community composition and size distribution of the herbivorous zooplankton. The impact of herbivory on phytoplankton depends on the relative abundance of certain herbivores with wide diets, high grazing rates, and rapid population growth rates. Population of these keystone herbivores are sensitive to fish predation. In addition, the size distribution of fishes and zooplankton and their migratory behavior largely determine the rate and spatial pattern of nutrient recycling in pelagic ecosystem. In whole-lake experiments, manipulations of fish community structure have caused significant changes in primary production, algal biomass, nutrient recycling, and sedimentation rates.

C. Relationships between Biodiversity and Ecosystem Stability

A major concern for limnologists is to predict response to stress. For a long time, the so-called conventional wisdom in ecology was that increased complexity within a community leads to increased stability. Complexity is used here to mean more species, more interactions between them, and more pathways. The basic assumption is that if the number of pathways increases, any blockage at one point of the network would be compensated for by the opening of another pathway. However, until now this conventional wisdom has not received much support from field or experimental work. Therefore some basic questions remain open and are of particular concern for freshwater lakes:

- How is system stability and resistance affected by species diversity, and to what extent could the integrity and sustainability of ecosystems be maintained in spite of species deletions resulting from degradation of environmental conditions?
- Are rare species an insurance of ecosystem stability? Do these rare species play a role in ecosystem functioning? One hypothesis is that the most stable ecosystems in terms of key functions are those richest in species. However, well documented studies of rare species substituting for declining common species in the maintenance of key freshwater ecosystem functions following disturbance are scant.

D. Role of Intra- and Interspecies Communication Systems in Ecosystem Dynamics

The structure of the biota is determined by complex interactions between individual organisms acting at different trophic levels. It is now becoming clear that besides energy transfer from one trophic level to another, there is an exchange of information between trophic levels through infochemicals. Moreover, sexual pheromones, but also sounds, electric signals, and visual communication play a role in shaping the structure of the biotic community. This diverse communication network has biological consequences and may modify the behavior of aquatic animals, such as migrations. This is a fairly new field of research.

E. Biological Productivity and Biodiversity: The Case of Eutrophication

The biological structure and internal biological control mechanisms of freshwater lakes are highly affected by lake water nutrient level and by the extent of nutrial loading. Limnologists distinguish oligotrophic lakes, which are generally deep with steep slope and are characterized by low nutrient levels and clear water. The biomass at all trophic levels is small. On the opposite, eutrophic lakes are often shallow with gradually sloping edges. The most characteristic features are high nutrients levels, the abundance of plankton, and low water clarity.

One concept of lake succession considers that lakes pass through different trophic states, beginning with low fertility or oligotrophy, gradually moving to a moderately productive or mesotrophic state to reach finally the eutrophic stage. This succession may happen in undisturbed lakes. However, eutrophication (sometimes called cultural eutrophication) is now widespread as a result of human activities. Eutrophication may be defined as the process of enrichment of a water body due to an increase in nutrient loading. The most important nutrients causing eutrophication are phosphates, nitrates, and ammonia. All these chemicals are abundant in waters released from sewage treatment works and from surface and groundwater runoffs in intensive agriculture areas.

The most obvious consequence of eutrophication is the increased aquatic plants and phytoplankton growth, an overall increase in biomass, and a shift in species composition of the lake. For example, at low P-concentrations, north European shallow freshwater lakes are usually in a clearwater stage; submerged macrophytes are abundant, potential piscivores are present in large numbers, and predation pressure on zooplankton is consequently low. At some higher P-concentrations, there is a shift to a turbid stage: submerged macrophytes disappear and the fish stock changes. The fish biomass rises and there is a shift from a system dominated by pike (*Esox lucius*) and perch (*Perca fluviatilis*) to one exclusively dominated by planktivorous-benthivorous fish, mainly bream (*Abramis brama*) and roach (*Rutilus rutilus*).

IV. THE CASE OF ANCIENT LAKES SPECIES FLOCKS

About a dozen lakes in the world are up to three orders of magnitude older than most others (Table III) (Mar-

TABLE III

Summary of Physical and Biological Characteristics of Some of the Larger Extant Ancient Lakes

Lake	Age (My)	Max. depth. (m)	Area (km^2)	Number of animal species	Number of endemic	Number of fish species	Number of fish endemic
Baikal	35–30	1,700	31,500	1,825	982	56	27
Tanganyika	9–12	1,470	32,600	1,470	632	330	241
Malawi	4.5–8.6	785	30.800			>600	>600
Victoria	0,6?	70	70,000			>500	>500
Titicaca	3	284	8,448	533	61	29	23
Ohrid	2–3	295	348			17	2
Biwa	4	104	674	600	54	57	11

In part from Martens, 1997.

tens, 1997). Such lakes have exceptionally high faunal diversity and levels of endemicity.

An important characteristic of ancient lakes biodiversity is the existence of "species-flocks." An aggregate of several species should be identified as a flock only if its members are endemic to the geographically circumscribed area under consideration and are each others' closest living relatives. Briefly speaking, a species flock has to be monophyletic. At present, different rich species flocks for fish and invertebrates have been identified in various ancient lakes, which are therefore exceptional natural sites for the study of speciation patterns.

The processes accounting for these radiations are a matter of debate, but there is more and more evidence that sympatric speciation may occur in isolated water bodies. These species flocks are sometimes considered to be a world heritage that is endangered and has to be preserve from destruction by human activities (Coulter *et al.*, 1986; Nagelkerke *et al.*, 1995) (see Table III).

A. Fish Species Flocks

The most striking feature of the Great East African Lakes (Victoria, Tanganyika, Malawi) is that each has its own highly endemic lacustrine Cichlid fauna. In Lake Victoria, according to our present knowledge, there is a cichlid species flock of more than 500 endemic haplochromine species. The true species number is almost certainly even higher (Seehausen, 1996). The age of this flock was estimated at 200,000 years, but it is most likely that Lake Victoria had entirely dried up as recently as 12,400 years ago, so that most of the endemic cichlid flock would have evolved during the past 12,400 years (Johnson *et al.*, 1996). In Lake Malawi, the diverse cichlid fauna of this lake could also total much more than 500 species. Species flocks are also reported for the clariid catfish *Dinotopterus* (10 species).

The Lake Tanganyika cichlids are slightly less diverse. However, morphological and electrophoretic data both suggest that several lineages of cichlids from Lake Tanganyika are much older than the Lakes Victoria and Malawi lineages and can be traced back to at least seven distinct ancestral lineages. Species flocks also occur in noncichlid families: seven Mastacembelid species, six species of the bagrid *Chrysichthys*, seven species of *Synodontis*, and four species of the Centropomid *Lates* (De Vos and Snoeks, 1994).

Rates of speciation in cichlids can be astonishingly fast. That has been known since the discovery of five endemic species of cichlids in Lake Nabugabo, a small lake less than 4000 years old and separated by a sandbar from lake Victoria. Still faster speciation rates were suggested by the finding that the southern end of Lake Malawi was dry only two centuries ago, while it is now inhabited by numerous endemic species and "color morphs" that are only found there and are believed to have originated during the past 200 years.

The remarkable diversity of the large barbs (genus *Barbus*) in Lake Tana (Ethiopia) constitutes a potential species flock that has been discovered recently. Nagelkerke *et al.* (1995) hypothesized that intralacustrine speciation has occurred among the barbs of Lake Tana and possibly is still going on.

In South America, the native fish fauna of Lake Titicaca includes the genera *Trichomycterus* and *Orestias*, both endemic to the Andean Altiplano. Twenty-four *Orestias* species are presently recognized in Lake Titicaca (Lauzanne, 1992). However it is probably not a monophyletic group but rather an assemblage that includes, in part, several species flocks.

Lake Baikal (East Siberia) hosts a very diverse fauna, with some 2500 described species (most of which are endemic), which might constitute 50% of the total amount. At present, Lake Baikal comprises 56 species and subspecies of fish, which belong to 14 families (Sideleva, 1994), a group of six species and subspecies, belonging to three families (Thymallidae, Coregonidae, and Acipenseridae), which are relatively endemic; another group of nonendemic fauna includes 21 species and subspecies, which belong to Cyprinidae, Percidae, Cobitidae, Esocidae, Gadidae, Salmonidae, Siluridae, and Eleotridae.

In Asia, Lake Lanao was formed by a lava flow that dammed the streams flowing southwest. The cyprinid fauna presents a widely acknowledged example of adaptive radiation. Of the 23 cyprinid species presently known from Mindanao Island, 15 are reported from Lake Lanao (Kottelat and Whitten, 1996). Unfortunately overexploitation and exotic introductions have decimated the fauna, so that now only few endemic cyprinids are still present.

Lake Biwa is the largest and oldest lake in Japan. The deep basin as seen presently is supposed to have been formed 300,000 years ago. Most endemic fishes exist also since that time (Kawanabe, 1996). Some 500 plant and 600 animal species have been recorded. At present there are 71 species and subspecies if freshwater fishes found in Lake Biwa and its tributaries. There are 13 endemic species and subspecies of fish in Lake Biwa.

The only more or less pristine species flocks left in Asia are to be found in the lakes the Malili River drainage of the Sulawesi Island (Celebes) in Indonesia (Kottelat & Whitten, 1996). Malili lakes (lakes Towuti, Matano, Mahalona, Wawontoa, and Masapi) constitute a system of lakes partially isolated from each other and completely isolated from other freshwaters. As a result, most of the animal species known from the lakes are endemic.

B. Invertebrates

Mollusks focus our attention on parts of the world that seem to be hot spots of endemicity, where the resident clades are remarkably more diverse than in other similar environments. Long-lived lakes are prime examples of these evolutionary theaters. Particular clades, such as the hydrobioid and ceratioidean prosobranchs and the planorbid pulmonates, show repeated patterns of diversification in both extant and fossil long-lived lakes, revealing the common patterns that make them prone to speciate (Michel, 1994). The process of diversification is tied to intrinsic characters shared by many of these clades: reproductive and dispersal strategies (brooders and poor dispersers), genetic structure (tightly constrained genetic systems), morphology (often relatively thick and ornamented shells), substrate specificity (hard bottom stenotopy), and physiology (depth tolerance). The most notable examples of these evolutionary theaters are extant lakes Tanganyika, Baikal, Ohrid, Titicaca, and fossil lakes Idaho, Biwa, and Turkana. However, examples of gastropod radiations are also found in river systems (see Table IV).

Among Crustacea, gammarids have also undergone an enormous evolutionary radiation in Lake Baikal, with a total of 259 species, 98% of which are endemic. There are also several species flocks of Ostracods reported from ancient lakes (Martens *et al.*, 1994).

V. MAJOR THREATS TO BIODIVERSITY IN LAKES

The concentration of people around freshwater systems has resulted in a much greater degree of degradation to these systems than most open marine or even terrestrial systems.

A. Competition for Water

Competition for water may result in the total or partial desiccation of lakes and ponds through various diversion and impoundment of tributaries. Water is withdrawn most often from aquatic systems for irrigation, flood control, and urban and industrial consumption. A spectacular example is provided with the Aral Sea, a large saline lake in the terminus of an extensive inland drainage basin in south-central Asia. Water diversion for irrigation purposes of most of the waters in inflowing rivers of the Aral sea, as well as poor agricultural practices, resulted in a marked fall of water level (c. 15 m) and an increase in salinity (from c. 10 to 30 g/l) since the 1960s. This changes have resulted in the degradation of the natural environment. Fish have virtually disappeared from the lake and the diversity of associated bird and wildlife communities has decreased. Many invertebrates also disappeared (Williams and Aladin, 1991).

B. Habitat Alteration

Siltation from erosion of the lake basin has direct adverse effects on fish by covering spawning sites, destroying benthic food sources, and reducing water clarity to

TABLE IV

Endemism in Ancient Mollusk Fauna (ancient lakes and rivers)

Lakes	Baikal	Ohrid	Tangan.	Titicaca	Biwa
Mussels	2	1	12	0	
% end.	0	0	75		
Sphaeridae	10	9	3	4	
% end.	30	22			
Prosobranchs	72	47	52	16	11
% end.	93	91	84	93	72
Pulmonates	61	25	16	3	16
% end.	77	48	6	33	62
Total species	145	82	81	19	27

visual feeding animals. However, the increase turbidity may have also indirect effects on biodiversity in lakes. Seehausen *et al.* (1997) provided evidence that increasing turbidity (as the consequence of deforestation and agricultural practices) by curbing the impact of sexual selection on sexual isolation is responsible for the decline in cichlid diversity in Lake Victoria. Actually, mate choice in these cichlids is determined on the basis of coloration, and strong assortive mating can quickly lead to sexual isolation of color morphs, which is increasing and probably started in the 1920s. By constraining color vision, turbidity interferes with mate choice (Seehausen *et al.*, 1997). The reduced effectiveness of signals causes relaxation of sexual selection for color, with consequent loss of male nuptial coloration and erosion of species diversity due to a breakdown of reproductive barriers. Dull fish coloration, few color morphs, and low species diversity are found in areas that have become turbid as a result of recent eutrophication. This is proof that human activities that increase turbidity destroy the mechanism of diversification and the maintenance of diversity.

C. Species Introductions

The introduction of alien fish into inland waters has occurred all around the world. The main goals of deliberate introductions by fishery officers were initially to improve sport fisheries and aquaculture, or to develop biological control of aquatic diseases, insects, and plants, or else to fill supposed "vacant niches" and improve wild stocks in old or newly created impoundments.

The introduction of alien species has been considered as the main causes of extinction of endemic species flocks in several ancient lakes. In Lake Lanao, the introduction of the white goby (*Glossogobius giurus*) in the early 1960s resulted in the elimination of numerous species of endemic cyprinid fish. In Lake Titicaca the rainbow trout *Salmo gairdneri* was accused of seriously threatening the endemic *Orestias* fauna and for having been responsible for the disappearance of species such as *Orestias cuvieri*. In Lake Biwa, the recent increases in numbers of the exotic bluegill *Lepomis macrochirus*, black-bass *Micropterus salmoides*, and *Channa maculata*, have been mirrored by serious declines in the native species *Onchorhynchus rhodurus rhodurus* (an endemic), *Hemigrammocypris rasborella*, and *Hymenophysa curta*.

Much has been said about the impact of the introduction of the Nile perch on the hundreds endemic haplochromines of Lake Victoria (Lévêque, 1997). In the early 1980s this impact was considered an ecological and conservation disaster (Coulter *et al.*, 1986). However, it was later recognized that predation by *Lates* may not be solely responsible for the depletion of haplochromine stocks, and that the haplochromine stock was already affected by fisheries before the establishment of *Lates*, particularly by unregulated fishing or by trawling techniques introduced in the Tanzanian part of the lake. Lake Victoria is now invaded by water hyacinth, and the remaining fish fauna is therefore more and more threatened.

Transport through ballast water is probably one of the most important pathways for alien species invasions in several places, including the North American Great Lakes (Mills *et al.*, 1993). That is the case for the zebra mussel introduced into the Great Lakes, apparently in 1985 or 1986, which spread dramatically throughout the waterways of both Canada and United States expansion with serious economical and ecological consequences. The recent finding of individual mitten crabs (*Eriocheir sinensis*), a European flounder (*Platichthys *flessus*), and the establishment of the alien gastropod *Potamopyrgus antipodarum* in Lake Ontario in 1995 demonstrate that the process of invasion is still going on at a fast rate.

One of the major problems in freshwater species introductions is their irreversibility, at least on scale of a human's lifetime. Once introduced and established, it is impossible, given current technology, to eradicate a fish, a mollusk, or a plant species from a large natural water body. As a consequence, we are likely to see a continued reduction in native aquatic biodiversity and an increased homogenization of the world's freshwater biotas.

D. Fisheries Practices

One of the major threats to the unique species flocks of ancient lakes are the fishing practices and particularly overfishing and introduction of new fishing practices. According to Coulter *et al.* (1986), the collective experience in recent years on the African Great Lakes seems to show that large-scale mechanized fishing is incompatible with the continued existence of the highly diverse cichlid communities. Cichlids appear especially vulnerable to unselective fishing because of their particular reproductive characteristics. The structure of endemic cichlid fish communities in the African Great Lakes can change dramatically within a few years when trawlers and other such fishing gear are used. Actually, a number of authors have recorded the effects of overfishing in Lake Victoria, from the decline of some species to the virtual disappearance of others, and the history of the fishery was briefly reviewed by Witte *et al.* (1992).

It has also been suggested that parks should be developed (Coulter *et al.*, 1986) and that fishing should be rendered impossible in certain areas by placing obstructions on the bottom that would snarl trawls. Lake Malawi National Park will very probably afford protection to widespread species, but no data are at present available to confirm this hypothesis. It is unknown yet whether these reserves can adequately preserve the integrity of populations, but that is probably only possible for stenotopic populations whose distribution coincides with the park area. The size of the reserves, the intensity of fishing in nearby areas, and the possible influence of pollution or introduced alien species should also be taken into account.

E. Pollution

Pollution can affect aquatic biota through direct mortality at any life stage or by sublethal effects influencing predation, foraging, and reproduction. The eutrophication of Lake Victoria during the past 25 years is quite well documented. Enhanced quantities of nutrients appear to have been entering this lake for many years, both through rivers and from aerosols as a result of human activities in its watersheds. The eutrophication could lead to increased oxygen demand in the lake's deep water and thus decrease the hypolimnetic volume habitable by fish during seasonal stratification. This phenomenon is partly responsible for the threatening or disappearance of cichlid species belonging to the haplochromine flock.

The release of sulfur and nitrous oxides from the burning of fossil fuels may be transported great distances before being transformed chemically into sulfuric and nitric acids and deposited as rain, snow, or dust. When acid rains occur over areas where waters are poorly buffered, the chemistry and biology of freshwaters can be changed dramatically. Many softwater lakes have been acidified both in North America and Europe, but evidence has accumulated for its occurrence in China, the former Soviet Union, and South America. Monitoring studies indicate a general impoverishment of species numbers in lakes as they become more acidic. Many lakes in the northeastern United States have lost 30% or more of the species in some taxonomic groups. In many northern European countries, acidification strongly modified the fish composition and abundance in lakes.

See Also the Following Articles

ENDANGERED FRESHWATER INVERTEBRATES • EUTROPHICATION AND OLIGOTROPHICATION • FISH, BIODIVERSITY OF • FRESHWATER ECOSYSTEMS • FRESHWATER ECOSYSTEMS, HUMAN IMPACT ON • INVERTEBRATES, FRESHWATER, OVERVIEW • RESOURCE EXPLOITATION, FISHERIES

Bibliography

Carpenter, S. R., and Kitchell, J. F. (1993). *The trophic cascade in lakes*. Cambridge University Press., Cambridge.

Coulter, G. W., Allanson, B. R., Bruton, M. N., Greenwood, P. H., Hart, R. C., Jackson, P. B. N., and Ribbink, A. J. (1986). Unique qualities and special problems of the African Great Lakes. *Environmental Biology of Fishes* 17, 161–184.

Crane, K., Hecker, B., and Golubev, V. (1991). Hydrothermal vents in Lake Baikal. *Nature* 350, 281.

Dodson, S. I. (1992). Predicting crustacean zooplankton species richness. *Limnology and Oceanography* 37, 848–856.

Dumont, H. (1994). On the diversity of the Cladocera in the tropics. *Hydrobiologia* 272, 27–38.

Fernando, C. H. (1980). The species and size composition of tropical freshwater zooplankton with special reference to the oriental region (Southeast Asia). *Int. Rev. gesamten Hydrobiol.* 65, 411–426.

Kawanabe, H. (1996). Asian Great Lakes, especially Lake Biwa. *Environmental Biology of Fishes* 47, 219–234.

Kottelat, M., and Whitten, T. (1996). *Freshwater biodiversity in Asia, with special reference to fish.* World Bank Technical Paper, no. 343.

Lauzanne, L. (1992). Native species. The *Orestias*. In *Lake Titicaca. A Synthesis of Limnological Knowledge* (C. Dejoux and A. Iltis, Eds.), pp. 405–419. Kluwer Academic Publishers, Dordrecht.

Lehman, J. T. (1988). Ecological principles affecting community structure and secondary production by zooplankton in marine and freshwater environments. *Limnol. Oceanogr.* 33, 931–945.

Lévêque, C. (1997). Biodiversity dynamics and conservation: The freshwater fish of tropical Africa. Cambridge University Press, Cambridge.

Martens, K. (1997). Speciation in ancient lakes. *Tree* 12(5), 177–182.

Martens, K., Goddeeris, B., and Coulet, G. (Eds.) (1994). *Speciation in ancient lakes.* Archiv für Hydrobiologie. *Advances in Limnology.*

May, R. (1994). Biological diversity: differences between land and sea. Philosophical *Transactions of the Royal Society of London*, B 343, 105–111.

Michel, E. (1994). Why snails radiate: A review of gastropod evolution in long lived lakes, both recent and fossil. *Arch. Hydrobiol. Beih. Ergebn. Limnol.* 44, 285–317.

Mills, E. L., Leach, J. H., Carlton, J. T., and Secor C. L. (1993). Exotic species in the Great Lakes: A history of biotic crises and anthropogenic introductions. *Journal of Great Lakes Research* 19, 1–54.

Nagelkerke, L. A. J., Mina, M. V., Wudneh, T., Sibbing, F. A., and Osse, V. (1995). Lake Tana, a unique fish fauna needs protection. *Bioscience* 45(11), 772–775.

Palmer, M. A., Covich, A. P., Finlay, B. J., Gibert, J., Hyde, K. D., Johnson, R. K., Kairesalo, T., Lake, S., Lovell, C. R., Naiman, R. J., Ricci, C., Sabater, F., and Strayer, D. (1997). Biodiversity and ecosystem processes in freshwater sediments. *Ambio* 26(8), 571–577.

Seehausen, O. (1996). Lake Victoria rock Cichlids—Taxonomy, ecology and distribution. *Verduyn Cichlids.*

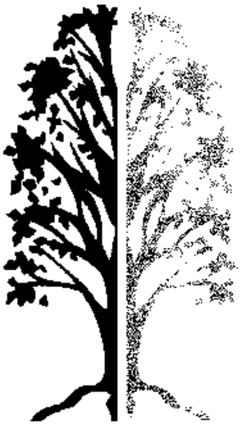
Seehausen, O., Jacques, J. M., and Witte, F. (1997). Cichlid fish diversity threatened by eutrophication that curbs sexual selection. *Science*, 277, 1808–1811.

Sideleva, V. G. (1994). Speciation of endemic Cottoidei in Lake Baikal. *Adv. Limnol.* 44, 441–450.

Williams, W. D., and Aladin, N. V. (1991). The Aral sea: recent limnological changes and their conservation significance. *Aquatic Conservation: Marine and Freshwater Ecosystems* 1, 3–23.

Witte, F. T., Goldschmidt, J., Wanink, M., van Oijen, K., Goudswaard, E., Witte-Maas, and Bouton N. (1992). The destruction of an endemic species flock: Quantitative data on the decline of the haplochromine cichlids of Lake Victoria. *Environmental Biology of Fishes* 34, 1–28.

Yu Sherbakov, D. (1999). Molecular phylogenetic studies on the origin of biodiversity in Lake Baikal. Tree 14, 92–95.



LANDSCAPE DIVERSITY

Debra P. C. Peters and Sarah C. Goslee
United States Department of Agriculture, Agricultural Research Service, Jornada Experimental Range

GLOSSARY

alpha landscape diversity Number and dominance of patch types within a landscape.

corridor Strip of land or water that differs from the adjacent landscape on both sides.

gamma landscape diversity Total number of patch types contained within a geographic region.

landscape Spatially heterogeneous area composed of a mosaic of interacting components (patches, corridors, and area of matrix).

matrix Background landform, habitat, or ecosystem in a landscape, characterized by extensive area, high connectivity, and major control over landscape dynamics.

patch Area that is relatively homogeneous with respect to the characteristics being examined, and that differs from its surroundings.

region Large geographic area that contains more than one landscape.

LANDSCAPE DIVERSITY IS A DESCRIPTION OF THE NUMBER AND DOMINANCE OF DIFFERENT PATCH TYPES contained within a spatially heterogeneous area. This article describes landscape diversity and its importance, as well as its relationship to species diversity.

I. INTRODUCTION

While biodiversity is usually considered at the species level, maintenance of biodiversity requires management at higher levels of organization, particularly at the landscape scale. It is difficult to manage for each threatened species individually. Alternatively, management can focus on the ecosystems that contain these species, and on the landscapes in which ecosystems are found. The relatively new discipline of landscape ecology provides insight into both landscape diversity and species diversity, and suggests a theoretical and practical basis for conservation planning.

There are three basic characteristics of landscapes that affect their diversity: structure, function, and dynamics. Structure is the most well-understood element of landscapes. It is also the most obvious—nearly any

Encyclopedia of Biodiversity, Volume 3

aerial view will show a mixture of different landforms, habitats, or vegetation types. The patch is the basic unit of landscape structure. The characteristics of patches and the spatial relationships among patches are important components of landscapes. The distributions of energy, materials, and species among patches differing in size, shape, abundance, and configuration are particularly important to patterns in diversity at the landscape scale. The other two elements of landscapes go beyond a description of spatial heterogeneity. Function is concerned with interactions among the spatial elements of a landscape, including flows of energy, materials, and species among patches. Landscape dynamics includes characteristics of both structure and function in order to examine changes in pattern and process over time. The conservation and management of biodiversity require understanding of all three elements, including the effects of human activities on the system. This article discusses each element in turn, and also considers the underlying determinants of landscape structure, including environmental heterogeneity and disturbance patterns. We then discuss classical and current issues in biodiversity management, and conclude with a case study of landscape diversity at the Sevilleta National Wildlife Refuge Long-Term Ecological Research site in central New Mexico, United States.

It is essential to keep the concept of "scale" in mind when considering landscape diversity. Spatial scale has two elements: grain and extent. Grain is the minimum resolution sampled, usually the cell size or quadrat size for ecological studies. The extent is the domain of the study, which is typically the size of the study area. Ecological processes often have characteristic spatial and temporal scales. This means that the grain and extent of sampling in both space and time may strongly affect the results of a study. For example, as quadrat size (grain) increases, species richness may increase, yet diversity of patch types within a landscape may decrease since fewer large quadrats can be found within a given area (Fig. 1a). As landscape size (extent) increases, more species and more patches of a constant size may be found that would increase both species and landscape diversity (Fig. 1b).

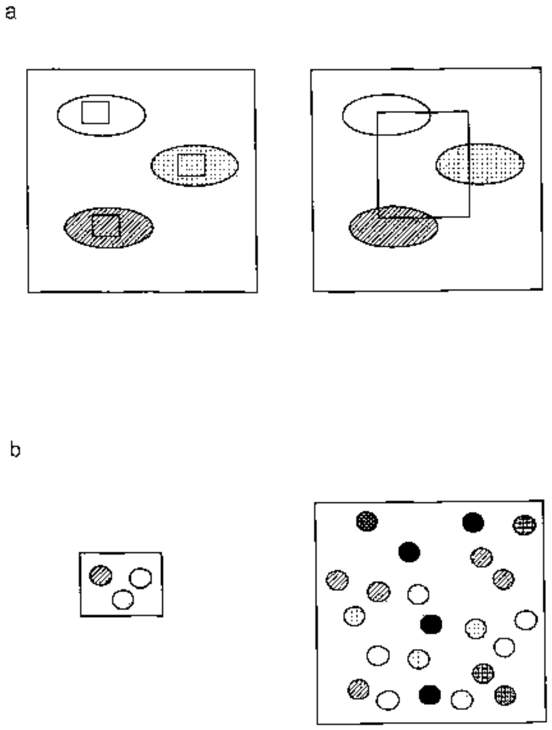


FIGURE 1 Landscape grain and extent: (a) as the grain or quadrat size of a landscape increases, yet the extent remains the same, the diversity of patch types decreases since fewer quadrats can be found in the same area; (b) as landscape extent or size increases, more species and more patch types of the same size can be found that result in higher landscape diversity.

II. DESCRIPTION OF LANDSCAPE STRUCTURE

Landscape structure can be most easily described at two hierarchical spatial levels, both of which are relevant to landscape diversity as well as to species diversity. At the lower level, the focus is on the attributes of individual patches, particularly size and shape. Description at the higher spatial level is concerned with the composition and pattern of the entire landscape and its mosaic of patches. The ability to quantify landscape structure at both levels allows the comparison of different landscapes. More importantly, interactions between landscape structure and function have implications for both species and landscape diversity.

A. Patch Description

A patch is a relatively homogeneous nonlinear area that differs from its surroundings. The definition and identification of individual patches and their boundaries are important steps in characterizing the structure of a landscape. In some systems, boundaries may be easily identified, such as between patches of agricultural field and adjacent woodland in human-dominated systems. In many cases, however, the boundary is not so clear, and patches are more difficult to delineate. Most methods of patch identification combine qualitative and quantitative approaches. A subjective determination of how different two areas must be in order for them to be considered separate patches is often needed. A number of quantitative techniques have been developed to group similar cells into homogeneous patches or to identify repeating patterns across a landscape (Turner and Gardner, 1990). Approaches such as blocking techniques, spectral analysis, and nearest-neighbor analysis are commonly used. Other techniques rely on the detection of edges or boundaries rather than identifying patches directly. These methods include moving window analysis and image analysis to characterize landscapes with sharp transitions.

Patch identification provides an excellent example of the importance of the spatial scale of the observer. From inside a forest, clumps of trees and grass-dominated openings appear to be separate patches with different vegetation and resource availability. From an aerial view, the entire forest appears to be a single patch. This illustrates the importance of the selection of spatial scale based on study objectives prior to the determination of patches and their edges.

Once the patches in a landscape have been identified, there are many ways to describe and quantify them (Riiters *et al.*, 1995). Only patch size and shape will be discussed here, since the relevance of these two attributes for species diversity is the most well understood. The relationship between patch size and species richness goes beyond the familiar species–area curve (Fig. 2). Although the number of species present in a patch tends to increase with patch size up to a certain limit, the kinds of species found also tend to vary with size. Two general types of species can be distinguished. Interior species are found primarily in the interior of large patches. These species often have very specific habitat requirements and are relatively rare. Migratory songbirds that are particularly sensitive to patch size and adversely affected by habitat fragmentation are interior species. By contrast, edge species are found near the edges of large patches and throughout small patches that consist mostly of edge habitats. Edge species are commonly occurring generalists that can use various habitat types, and are often introduced species. Because small patches consist mostly of edge with little interior area, they often have the highest species densities, but contain few or no rare species. Large patches, on the other hand, are mostly interior area with lower species densities per unit area, but they contain more rare species and a higher total number of species.

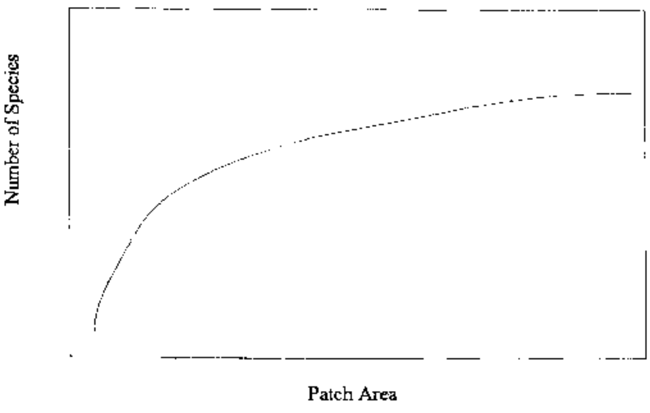


FIGURE 2 Species–area curve showing the increase in species number with increasing patch size tending toward a regional limit.

In an important study of tropical deforestation in the Amazon rain forest, species in patches of various sizes were compared to evaluate the importance of patch size to species number (Lovejoy *et al.*, 1984, 1986). Large patches were richest in species and small patches were found to contain only edge conditions. Patch size had important effects on different species, including trees, insects, birds, and mammals, that were noticeable in a short time. This study is one of the few where patch size was experimentally manipulated to allow comparison with pretreatment conditions as well as control patches.

A simple measure of patch shape is the perimeter : area ratio. This measure is often standardized so that the most compact possible form, either a square or circle, is equal to 1. More complex shapes have increasingly higher numbers. Another common index of shape complexity is the fractal dimension, which is also derived from the perimeter and area of a patch. The fractal dimension of a patch is between 1 and 2; a simple shape will have a lower fractal dimension than a more complex shape. Figure 3 illustrates the amount of interior area available in patches of different shapes. Both patches have an area of 25, but the perimeter of shape a is 20 while the perimeter of shape b is 32. Using a scaled perimeter : area ratio, a has a value of 1 and b has a value of 1.6. Assuming that interior area is at least 1 unit from any patch edge, a has an interior area of 9, but b has an interior area of only 2. Thus, a is more compact and less convoluted than b, where more of the area is closer to its edge and can interact with the area surrounding the patch. This suggests that the overall flow of species and resources between b and its surroundings is higher than that between a and its surroundings. It is also expected that a would have higher richness of interior species than b, which would have higher richness of edge species.

B. Landscape Description

At the landscape level, there are two basic components of structure: composition and pattern. Composition refers to the parts (i.e., patch types) that make up the landscape, and pattern refers to how these patches are arranged. Though these two components are conceptually different, in practice they are often related. For example, the pattern of agricultural fields on a landscape is likely to be different from the pattern of undisturbed woodland.

Landscape composition can be measured in ways analogous to measurements of species composition (Romme, 1982). The most straightforward approach is landscape richness or the number of different patch types in a landscape. Another approach includes the relative abundance or dominance of different patch types along with richness. These landscape indices were derived from information theory and are closely related to species diversity measures, such as the Shannon–Wiener and Simpson indices, which are used to describe alpha species diversity (Turner, 1989; Huston, 1994). Using one of these indices, a landscape containing many small patches of different types would have a higher diversity value than a landscape consisting of one large patch and several smaller patches, even if the total number of patches is the same for both landscapes. Landscape measures of richness and evenness were used in a study conducted in different patch types in Yellowstone National Park (Romme, 1982). Changes in landscape diversity through time were related to fire frequency, and were hypothesized to have important effects on species diversity as well as wildlife habitat (Romme and Knight, 1982).

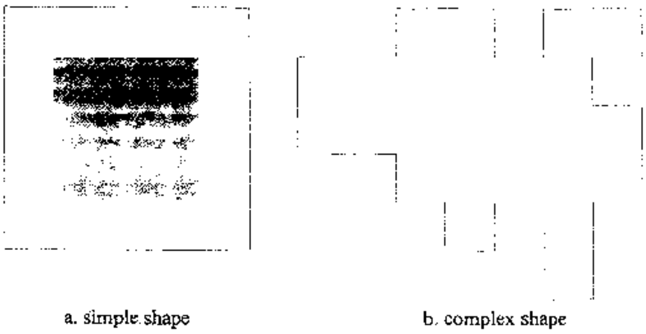


FIGURE 3 A simple shape and a more complex shape of equal area. The interior area is shaded.

Measurements of landscape diversity are analogous to common measurements of species diversity (Whittaker, 1960, 1972). Alpha species diversity is a measurement of species richness (number) and evenness (dominance or distribution) within a patch. Similarly, alpha landscape diversity is a measure of the number of patch types in a region (O'Neill *et al.*, 1988). Large-scale species diversity is called gamma diversity. Gamma landscape diversity of ecosystems is sometimes called ecodiversity (Rowe, 1992; Lapin and Barnes, 1995). The third form of species diversity, beta diversity, describes species turnover along a gradient. Beta diversity has no analog at the landscape level, but is sometimes estimated as gamma diversity/alpha diversity, which gives an average regional beta diversity.

Because different patch types provide different habitats and species compositions, one might expect that the total number of species in a landscape would increase as landscape richness increases. This idea was supported by a study that compared plant species richness in Rhode Island Audubon refuges varying in terrain and soil properties (geomorphological measures) (Nichols *et al.*, 1998). In a related study of one landscape, the diversity of trees and shrubs was higher on plots with the greatest geomorphological heterogeneity, indicating an important connection between landscape diversity and species diversity (Burnett *et al.*, 1998). Although this simple relationship between landscape and species diversity is generally true, the interactions between landscape composition and species diversity are more complex, in part because of species preferences to edge or interior types of habitats. The species found in a diverse landscape with many small patches are mostly edge species. Interior species are found only in landscapes with large patches, even though these landscapes have a lower diversity. The total number of interior species increases with the number of large patches on a landscape, similar to the species–area relationship in Fig. 2. Thus, the type of species that increases with increasing landscape diversity depends on the change in the size and configuration of patches within the landscape.

Landscape pattern, or the spatial arrangement of patches, can be measured in a number of ways, some of which are extensions of the patch-level metrics already discussed. These measures focus on patch abundance without regard to location in the landscape. The distribution of patch sizes can be determined within a landscape and used as information in the management of habitat patches for species that are sensitive to patch size or spatial arrangement, such as the spotted owl in the Pacific Northwest. The effects of forest clear-cutting on changes in patch structure and implications for interior and edge species provide another example of the importance of these measures. Shape complexities and boundaries can also be scaled up from the patch to the landscape level using the fractal dimension and perimeter : area ratio (Milne, 1988).

The second type of measurement for landscape pattern explicitly considers the location of patches relative to each other and includes patch abundance as well. Dispersion indicates the tendency of patches of one type to be distributed either uniformly, randomly, or aggregated. Contagion describes the tendency of patches of two different types to be near each other. Connectedness can be quantified using nearest-neighbor probabilities that reflect the degree of fragmentation in the landscape. All three of these indices have implications for the flow of species and resources between patches of the same and different types, and thus have important effects on species diversity.

Two additional structural elements other than patches may be recognized in many landscapes. The second element is the matrix or the background landform, habitat, or ecosystem in a landscape. The matrix is characterized by extensive cover, high connectivity, and major control over landscape dynamics. Forest patches contained within a matrix of subdivisions are functionally very different from forest patches surrounded by agricultural land. Corridors, strips that differ from the adjacent landscape on both sides, are the third element of landscapes. Corridors are usually linear and always highly connected; stream networks and roadways are common examples. Corridors may also connect larger patches of a similar type, such as a stream flowing between two lakes.

The patch–matrix–corridor model of landscape structure is illustrated in Fig. 4. Corridors may be particularly important for preserving species diversity by allowing movements of species across diverse landscapes. Corridors can also adversely affect species diversity by allowing nonnative or exotic species to invade and reduce the number of native species in an area. An example is the extensive spread of cheatgrass, an annual introduced to North America in shipments of grain from Asia and Europe in the 1880s (Mack, 1981). Movement of cheatgrass seed along railroad and cattle trail corridors in the early 1900s spread this grass throughout much of the northwestern United States, resulting in changes in species composition and dominance as well as losses of diversity.

III. CONTROLS ON LANDSCAPE DIVERSITY

Heterogeneity or diversity of landscape structure arises from a number of factors. Patches can be produced

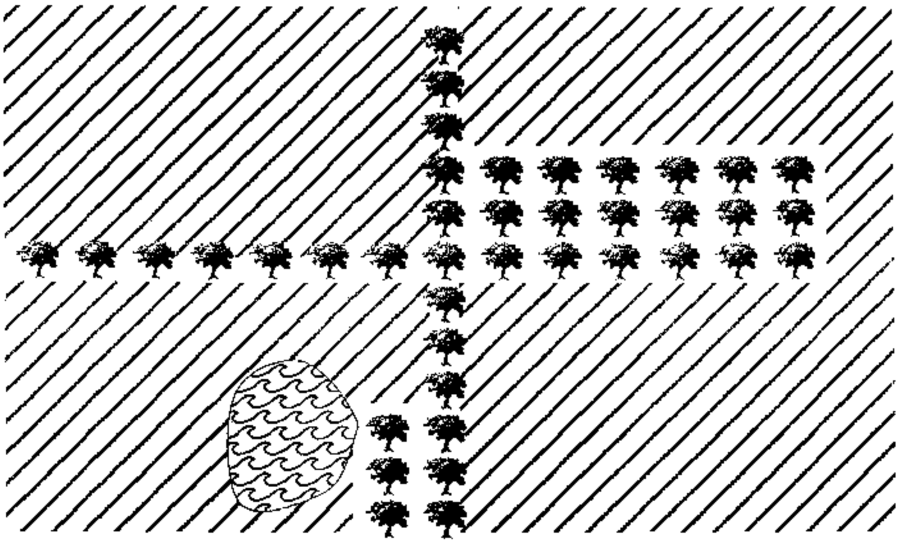


FIGURE 4 A matrix of agricultural land with patches of two types: woodlots and a farm pond. Windbreaks of trees act as corridors connecting the woodlots. Hedgerows and wooded windbreaks often create travel routes for wildlife.

through biotic or abiotic causes, including natural- or human-caused disturbance, fragmentation, regeneration, and persistent differences in environmental resources. Once a patch is formed, environmental conditions or interactions among organisms may change through time, leading to successional dynamics on the patch. A landscape consisting of patches in various successional stages is called a "shifting mosaic" (Bormann and Likens, 1979). The spatial pattern of patch formation and the changes within patches are collectively called "patch dynamics" (Pickett and White, 1985). The patch dynamic mosaic is part of the broader landscape transformation that includes changes in corridors and the matrix, as well as in the dynamics of species and ecosystem processes. These dynamics are discussed in Section V.

Biotic causes of patch generation include the local dispersal of seeds into a landscape, such as by an invasive weed, and the spatial segregation of populations or communities as a result of competition. Spatial structure can also be generated by differences between species in their dispersal abilities and rates of mortality. Naturally occurring and human-created disturbances are common causes of patch formation. A wide variety of natural disturbances are possible, including mud slides, avalanches, windstorms, ice storms, herbivore outbreaks, animal grazing, trampling, and digging, as well as fire. Mounds produced by badger digging activities in tallgrass prairie are an example of patch-producing disturbances that have important influences on patch structure as well as species diversity (Platt, 1975). Human activities, such as forest cutting, altered fire regimes, cultivation, urban development, introduction of pests, and strip mining for surface coal and minerals, also produce disturbance patches. Many landscapes are influenced by both natural- and human-caused disturbances, and distinguishing the separate effects on landscape diversity can be difficult. In a recent study, pollen and charcoal were collected from small lakes in Massachusetts in order to reconstruct long-term vegetation dynamics as related to disturbance history (Fuller *et al.*, 1998). This reconstruction over the past 1000 years included the period of time prior to European settlement when the primary disturbances were fire and wind. Landscape patterns in forest composition following settlement by Europeans were largely influenced by clearing of forests for agricultural purposes and timber. These researchers found that the past history of disturbance as a result of settlement has persistent effects on current landscape patterns.

Landscape fragmentation is closely related to disturbance. Many forms of disturbance effectively break up large patches into smaller pieces. Decreases in patch size, connections between patches, and total interior area as a result of fragmentation have important implications for species and landscape diversity. As landscapes become more fragmented, patch diversity increases with subsequent increases in edge species, exotic species, and generalists. Richness in interior species tends to decrease. Fragmentation of landscapes by human activities is considered a major threat to biodiversity worldwide (Saunders *et al.*, 1991; Bierregaard *et al.*, 1992). A major focus of the field of conservation biology is the design of nature reserves to maximize the likelihood of species existence and to minimize the loss of species to extinction. These processes are discussed in Section VI.

Another cause of patch formation is environmental heterogeneity, which refers to variation in soils, topography, and other landform features. This variation in the physical environmental leads to heterogeneous or patchy spatial distribution of resources, including water, nutrients, and light. Plant species found in a resource patch can differ from species in other patches containing different levels of resources. The importance of spatial heterogenity to species diversity has been well documented, and is most closely related to beta species diversity. The extension of these ideas to landscape diversity has occurred more recently, and several studies have linked measures of alpha and beta species diversity with landscape diversity (Romme, 1982; Lapin and Barnes, 1995). Large-scale gradients in landscape diversity can also be related to broad-scale patterns in the envionment. For example, spatial variation in climate, topography, and soils was found to be strongly related to latitudinal gradients in richness of land cover types across the continental United States (Wickham *et al.*, 1995).

IV. LANDSCAPE FUNCTION

Interactions among the spatial elements of a landscape are the major components of landscape function. These flows of energy, materials, and species among patches, or among patches, corridors, and the surrounding matrix, are at least as important to the maintenance of diversity as patch size and configuration. However, these flows have not been as well studied as landscape structure. An example of flows among different patch types is the dispersal of seeds from forest patches into clear-cuts, which has important effects on vegetation dynamics in these open areas. Boundaries or edges between patches or between patches and the mosaic often control the strength of interactions or the amount and kinds of materials that can move between the landscape elements. Because of the importance of edges, boundaries of patches can have very different characteristics than interiors. For example, edges of recently disturbed tropical forest patches have greater tree mortality and increased recruitment of early-successional species compared to interior areas (Bierregaard *et al.*, 1992). Boundaries can also change location through time with resulting effects on landscape structure. In contrast to boundaries, where movement is generally restricted, corridors linking similar landscape elements tend to improve or enhance flows. Movements of organisms through corridors become increasingly important as the landscape becomes more fragmented (Saunders and Hobbs, 1991).

There have been several studies of the influences of landscape structure on flows of organisms and materials. Patchy environments in Yellowstone National Park were found to be more resistant to large fires than were homogeneous landscapes, and after burning they had a greater ability to maintain water quality (Knight and Wallace, 1989). Historical migration patterns of wild ungulates, such as wildebeest in the Serengeti of Africa and bison in North America, were mostly related to patterns of rainfall that were spatially variable both locally and regionally (reviewed in Coughenour, 1991). Changes in landscape structure through fencing and urbanization have restricted migration patterns and resulted in animal overabundance and overgrazing in wildlife preserves. Biogeochemical fluxes, such as CO_2 and various forms of nitrogen, can also be affected by patches within a mosaic structure created by human land use. Gene flow and metapopulation dynamics are other examples of processes that respond to spatial structure in a landscape. A population that is spatially subdivided into patches that are connected through dispersal is called a metapopulation. Movement of individuals between subpopulations can reduce the risk of local extinction of species within small isolated patches.

V. LANDSCAPE DYNAMICS

Landscape structure and function can change for many reasons and in many ways. Changes can happen over very small or very large areas, and over short or long time spans. The gap caused by a single tree falling in the forest during a storm is small and temporary, while an entire forest may be leveled by a hurricane and take decades to centuries to recover.

Vulnerability or sensitivity to change varies from landscape to landscape. This vulnerability (or, conversely, stability) is traditionally divided into two components: resistance and resilience. Resistance is the ability of a patch or landscape to remain unaffected by a disturbance. A grassland is much more resistant to wind damage than a forest, since grasses can bend with the wind without breaking. Resilience is the ability of a patch or landscape to recover after a disturbance. Temperate forests recover after clearing much more quickly than tropical forests (which may never recover) owing to differences in soil depth and fertility.

Change in a patch or landscape can be caused by any number of factors. Some of these are intrinsic to the population being studied, including recruitment, growth, mortality, and spread or migration, which can lead to invasions or extinctions as well as changes in patch boundaries. Other causes are extrinsic to the ecosystem and are imposed by outside forces, such as climate change and disturbance events. Human transformations of the landscape include deforestation and reforestation, urbanization, corridor construction, and agricultural conversion. The effects of consumer, pathogens, and especially people can be considered either intrinsic or extrinsic depending on the particular point of view. The potential causes of change may be interrelated in complex ways. A drought may make a forest more vulnerable to pathogens, or a new clearing may increase the vulnerability of adjacent trees to windthrow.

Changes in landscape structure can have several spatial and temporal forms. Patches can shrink or expand, or be lost entirely. Successional dynamics on patches can lead to a shifting mosaic of patch types through time. Species interactions with other species and with their environment, as well as dispersal of new species into patches, are primary determinats of the regrowth of plants on these successional patches. Changes in patch size and shape can occur along edges, such as the clearing of forest to increase the size of a cultivated field (Fig. 5a). A new patch type may spread outward from a corridor (Fig. 5b). For example, housing developments often spread from the course of new roads. Alternatively, a patch type may spread out from a nucleus that could be a remnant of a previous vegetation type, or an introduction site for a new patch type (Fig. 5c). Some changes are nearly instantaneous and occur over very short periods of time, such as the effect of fire. Other changes occur slowly and take a longer period of time to develop, such as suburbanization and desertification.

Patch configuration on a landscape can also change. Patches can become perforated by other patch types,

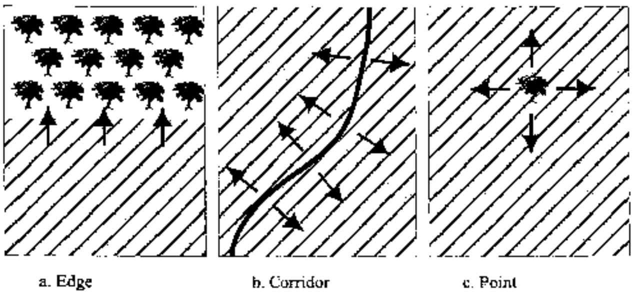


FIGURE 5 Three common patterns of landscape change. Arrows indicate the direction of spread. (a) An edge where agricultural land encroaches on forest along a field boundary. (b) A corridor where development spreads outward from a new road. (c) A point where reforestation proceeds outward from an isolated forest patch.

TABLE I

Transition Matrix for a Hypothetical Landscape with Three Land Use Types[a]

	Forest	Agriculture	Developed
Forest	60	25	15
Agriculture	10	75	15
Developed	0	2	98

[a] Each cell shows the percentage of the landscape area that changed from the patch type in that row to the patch type in that column over a single time step. For example, 25% of the original forest land was cleared for agriculture and 15% was developed. Figure 6 shows this projected change over 25 time steps.

and large patches can be fragmented into several smaller patches. Landscape fragmentation, particularly in the tropics, is having severe effects on species biodiversity. Some of the potential consequences of fragmentation include the loss of patch types and their characteristic species, decreased connectivity with its repercussions for species movements, and decreased interior area. The biggest consequence for species diversity is the associated loss of interior species and the increase of generalist or edge species.

Landscape-level dynamics are often studied with ecological models since the temporal scales of interest are often greater than the human life span, and experiments are difficult to perform at large spatial scales. There are four general classes of models that are used to predict landscape dynamics: transition probability models, individual-based models, ecosystem process models, and biogeographic models. Transition process models are useful when the factors causing landscape change are not represented mechanistically. For example, assume that a landscape with three patch types was sampled twice, before and after an event. A table can be constructed showing the percentage of each patch type that stayed the same or that was transformed into a different patch type in this hypothetical landscape (Table I). Each row shows the fate of a particular patch type. Over a single time step, 60% of the forest land remained forested, 25% was converted to agricultural uses, and 15% was developed. These transition probabilities can be used to extrapolate into the future by individual time steps. Figure 6 shows the projected change over 25 time steps if this original landscape had 50 units of forest, 25 units of agriculture, and 10 units of developed land. This landscape will stabilize with a high proportion of developed land and a very small forest area. Transition analyses are very simple to conduct and can be useful for examining the effects of various probabilities and initial conditions. In the simple form presented here, no allowance was made for variations in the rate of change, and no specific spatial component was included.

Individual-based simulation models are useful when information is known about the mechanisms underlying changes in landscape structure. These models incorporate life-history traits of individuals and the mechanisms by which they interact with their environment in order to predict landscape-level dynamics. Landscapes are simulated by linking plots together in a grid or transect. Plots are spatially interactive through processes such as seed dispersal. Spatially interactive individual-based models can represent a variety of environmental conditions, including differences in soil properties, climate, and disturbance regime (Coffin and Lauenroth, 1994). These models are most commonly used for evaluating changes in the diversity of groups of similar species (i.e., functional types) rather than species diversity.

A third class of models simulates ecological processes, including rates of nutrient cycling, water balance, and primary production. These models have been linked with geographic information systems (GIS) to simulate large regions. Effects of climate, soil texture, and management on soil organic carbon dynamics were simulated for the central grasslands of the United States (Burke *et al.*, 1991). Across this large region, soil organic carbon increased with precipitation and decreased with temperature and percentage sand content. Biogeographic models are a fourth class of models that can be used to investigate vegetation responses to environmental heterogeneity. These models incorporate large-scale variations in climate and soils, as well as water

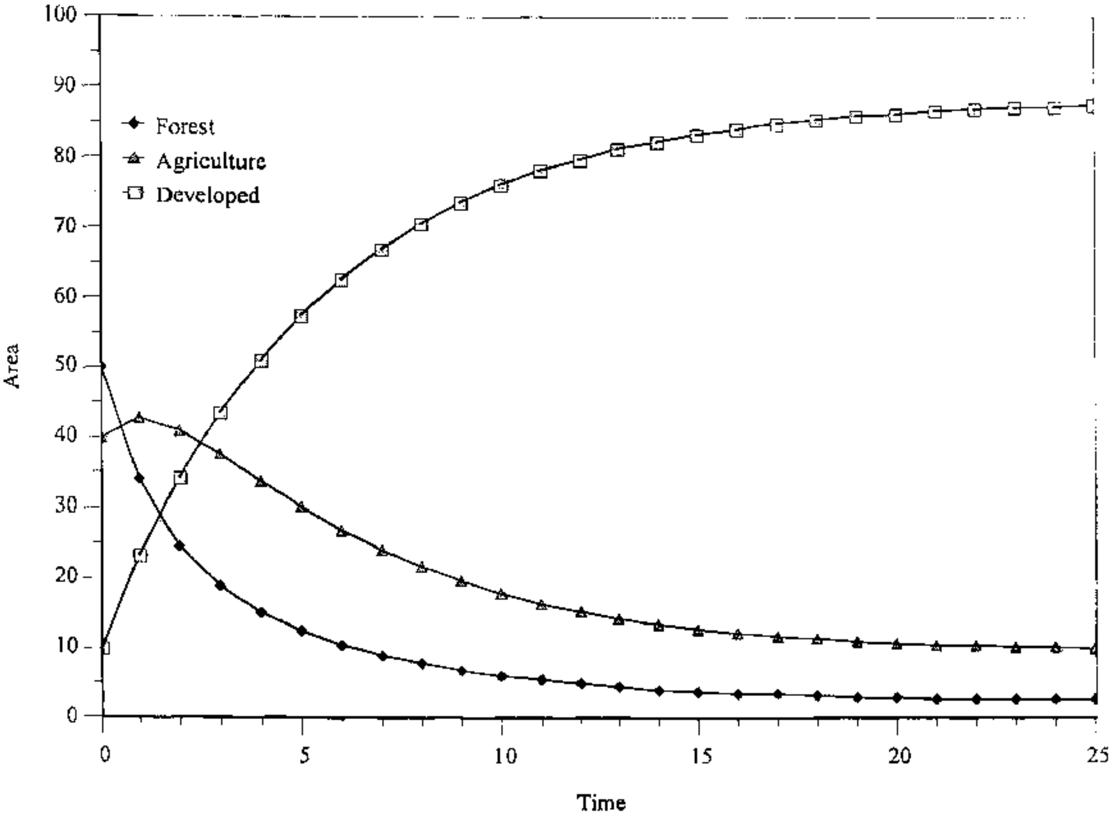


FIGURE 6 Transition matrix analysis of a hypothetical landscape with three patch types of differing abundance: forest, 50 units; agriculture, 25 units; and developed, 10 units. Transition probabilities are shown in Table I.

and energy constraints on plant growth, to simulate continental and global patterns in vegetation. Biogeographic models are most useful for simulating responses of plant functional types at large spatial scales, to either equilibrium or transient environmental conditions (Prentice *et al.*, 1992; Neilson and Drapek, 1998). These models do not explicitly incorporate landscape-scale processes or flows among patches.

Although each of these types of models has traditionally been used independently, the linking of different models together is a recent development that has considerable potential for addressing issues related to landscape diversity. Because of important feedbacks between species and rates of ecosystem processes (Schulze and Mooney, 1994), linking a spatially interactive individual-based model with an ecosystem model has a large potential for simulating the dynamics of landscape structure and function as well as changes in species diversity. An important first step was illustrated by connecting a nonspatial individual-based model with a nutrient cycling model to examine the importance of soil heterogeneity to forest responses to global climate change (Pastor and Post, 1988). The incorporation of landscape-scale flows of water, carbon, and nutrients into a spatially interactive individual-based model is an important research area for predicting landscape diversity dynamics. Recent linkages between biogeochemical and biogeographic models are another area of potential application to landscape diversity, especially if the plant functional types become more resolved and landscape-scale processes, such as disturbance regime, are included.

VI. BIODIVERSITY PLANNING AT THE LANDSCAPE LEVEL

To preserve species diversity most effectively, management plans must preserve the habitats and landscape structures needed by the target species, rather than simply preserving the species in isolation from the larger, potentially changing environment. Management practices aimed directly at a particular species run the risk of losing ecosystem functions that might actually be crucial for the target species, but that were unknown when the management plan was created. Furthermore, maximizing benefits for one species may threaten others. The ideal is to preserve overall ecosystem health, including species diversity. Unfortunately, this is easier said than done. Much of the effort of conservation biologists has been directed toward learning how to manage ecosystems, at both theoretical and practical levels.

One of the classic debates in conservation biology centers around the "best" reserve design. If limited resources are available to purchase land, is it better to establish one big reserve or a few smaller ones? This has become such a famous and controversial issue that it has its own acronym, SLOSS (single large or several small). A large reserve provides the most potential habitat for interior species, which are usually the ones most in need of protection. However, a single reserve is vulnerable to all sorts of disasters. If a major hurricane or a pathogen hits that reserve, there are no other reserves to take its place. The establishment of several smaller reserves minimizes the risk of losing everything at the same time. However, a minimum size is needed to sustain populations of interior species as well as to preserve the characteristic species diversity and species composition of the ecosystem. Furthermore, reserves do not operate like isolated islands, thus connections between reserves and the surrounding habitats are also important.

A related concept involved in determining the optimum size of a nature reserve is the minimum dynamic area (Pickett and Thompson, 1978). Assuming the disturbance regime of an area is known, the frequency, areal extent, and recovery time can be used to determine the smallest reserve area in which there will always be some mature patch types to provide a species source for the rest of the area as it recovers from disturbance. If a patch is smaller than this minimum dynamic area, it will likely be eliminated through time simply from natural disturbances.

Given the large number of species on the planet, it is impossible, or at best impractical, to manage for every one of them. Instead, conservation biologists are now trying to identify ways to simplify the task of landscape-level management. The most promising methods identify one or a few important species, and concentrate on their management. One tactic is to manage keystone species, those on which important ecosystem functions or other species depend. Another approach is to target umbrella species, those with large ranges or broad habitat requirements. Managing for these species will automatically save many other species with smaller or less inclusive requirments. A similar method identifies a set of focal species, each of which is sensitive to a particular aspect of landscape structure or function. One of the focal species might be especially vulnerable to habitat fragmentation, whereas another might require a high level of connectivity. The protection of this set of sensitive species provides the management goals. When the requirements of the sensitive species are met, other species will be provided for as well.

VII. CASE STUDY: SEVILLETA NATIONAL WILDLIFE REFUGE

The Sevilleta National Wildlife Refuge (SNWR; 34.5°N, 106.9°W), located approximately 75 km south of Albuquerque, New Mexico, provides an excellent example of landscape diversity and its relationship to species diversity. This 100,000-ha wildlife refuge was established in 1973 and is currently managed by the United States Fish and Wildlife Service. The refuge is also a Long-Term Ecological Research site funded by the U.S. National Science Foundation. The climate at the SNWR is semiarid to arid, with low amounts of precipitation and high temperatures during the April to October growing season. Mean annual precipitation over the past 65 years was 23.4 cm. (sd = 70.4 cm) and average annual temperature was 14.1°C (sd = 0.7°C).

The SNWR is uniquely located at the ecotonal boundary between four major grassland–shrubland biomes found within the continental United States. Two of these biomes, shortgrass steppe ecosystems and Chihuahuan desert grasslands, form transition zones in the eastern part of the refuge (Figs. 7a and 7b). Patches of variable size (<10 m^2 to >1000 m^2) and shape occur and result in high landscape diversity (Fig. 7b). These patches can be differentiated into one of two patch types based on the cover of the dominant plant species (Gosz, 1995; Kröel-Dulay *et al.*, 1997). The vegetation of some patches consists mostly of blue grama (*Bouteloua gracilis*), the dominant species in shortgrass steppe ecosystems. A second patch type occurs where the majority of cover is black grama (*Bouteloua eriopoda*), a dominant grass in Chihuahuan desert ecosystems. Species richness is similar for both patch types, although the cover of plants is higher in black grama compared to blue grama patches (Coffin, 1997). A transect across the conceptual landscape shown in Fig. 7b goes through each patch type as well as the matrix vegetation where similar cover of both species occurs (Fig. 7c).

Within each patch, a smaller scale of heterogeneity also exists due to disturbances associated with the burrowing activities of bannertail kangaroo rats (see Fig.

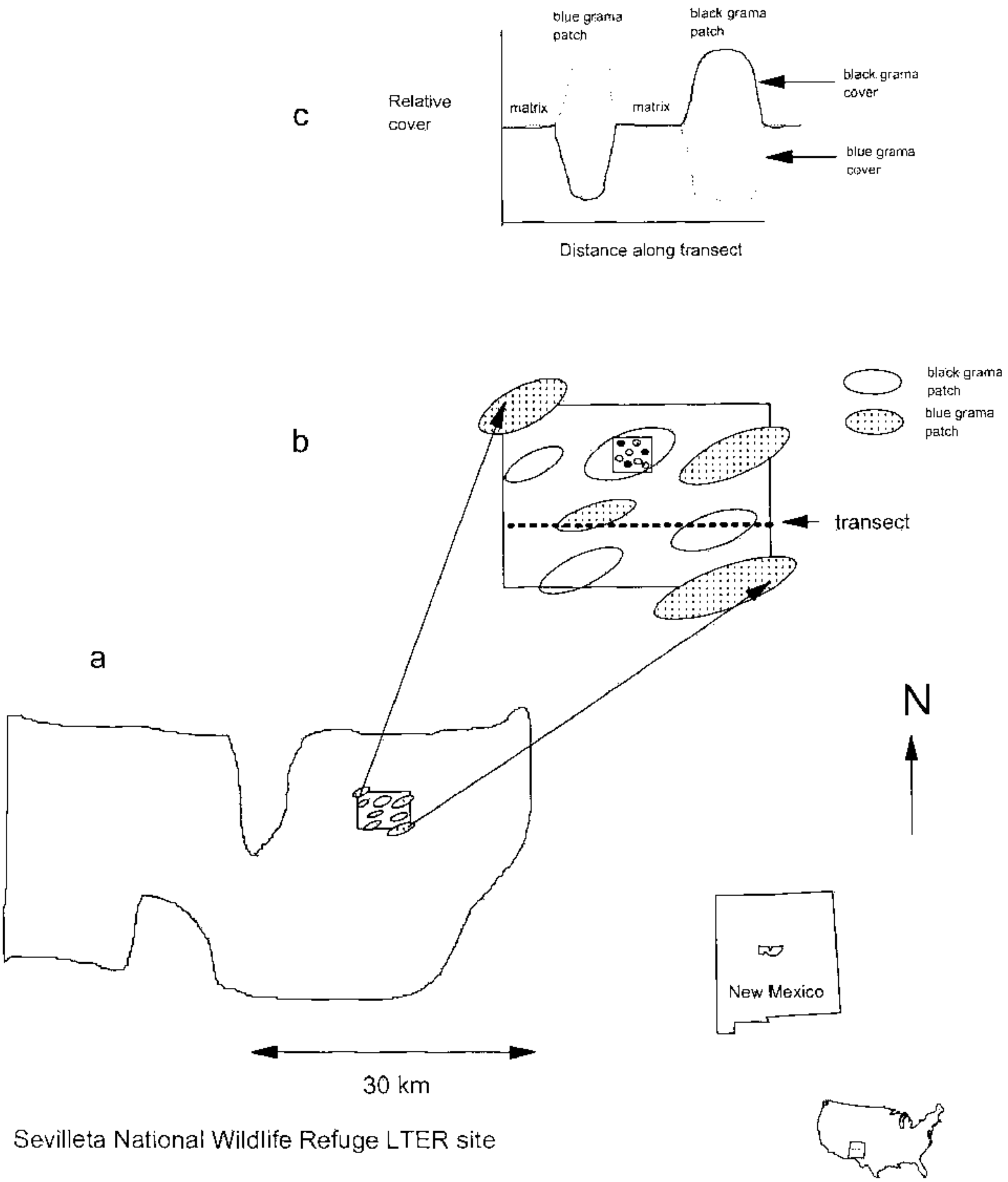


FIGURE 7 Patch types at the Sevilleta National Wildlife Refuge LTER site (SNWR). (a) Location of the SNWR in central New Mexico, United States, and location of patch types in the eastern part of the refuge. (b) Location of two patch types on the landscape. The transect across the landscape goes through patches dominated by either blue grama, black grama, or the matrix containing both species. (c) Cover of blue grama and black grama along the transect, showing one approach to identifying and delineating patches.

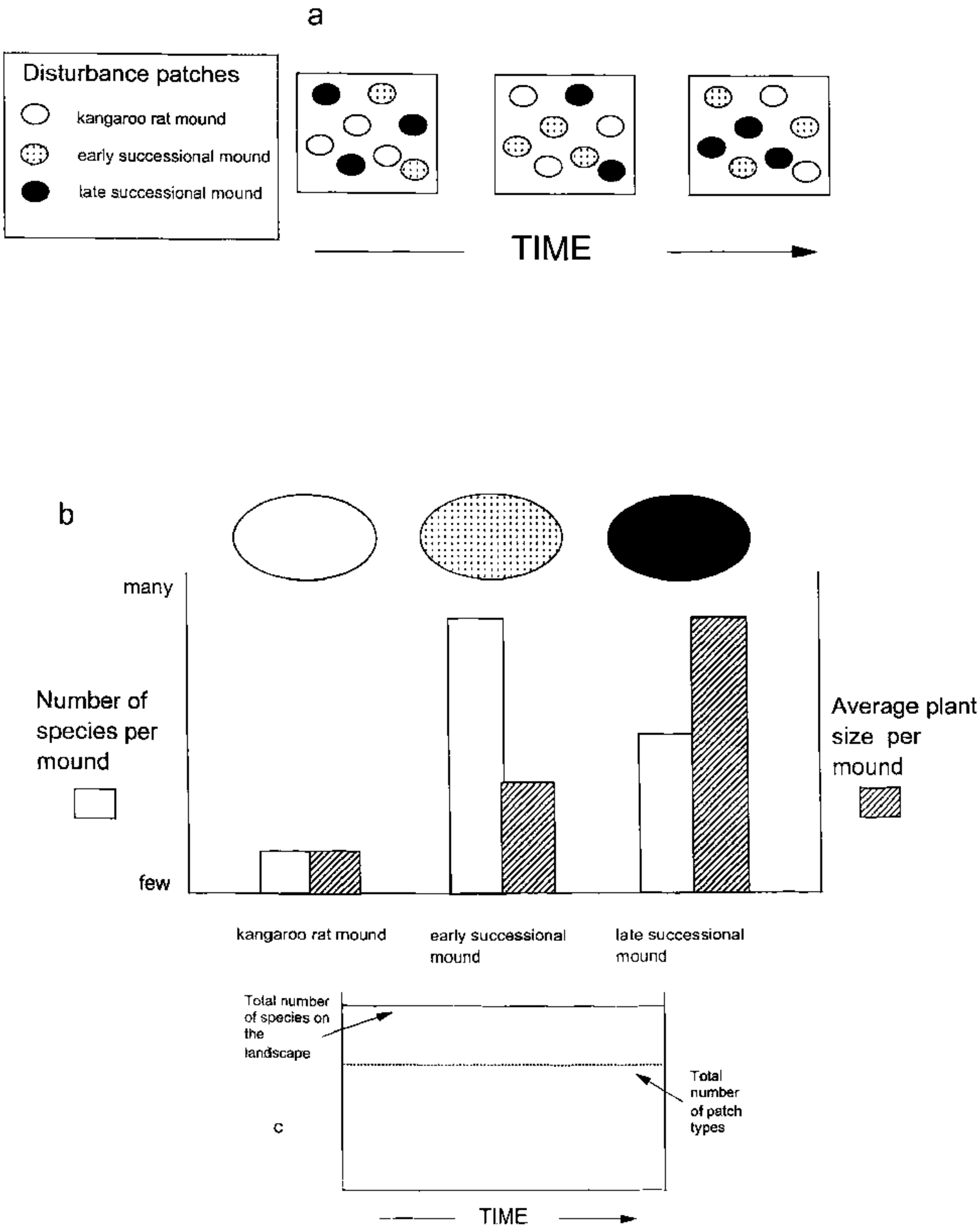


FIGURE 8 Small-scale heterogeneity within one black grama patch from Fig. 7b. Three types of mounds can be defined based on plant species composition and burrowing activities of bannertail kangaroo rats. (a) Mounds change from one type to another through time as kangaroo rats abandon some mounds and invade others. (b) Number of plant species and average plant size for each mound type. (c) Because the number of mounds within each of the three types is constant in this hypothetical landscape, the total number of species found on the landscape is also constant.

7b). Mounds can be distinguished into one of three types based on plant species diversity as well as composition (Fig. 8a). Active mounds are the site of frequent burrowing, thus only plant species well adapted to disturbance can survive there. Typically this vegetation consists of small plants representing few species (Fig. 8b). After mounds are abandoned and burrowing activities cease, more plant species can survive to larger sizes on these early-successional mounds. Through time, competition among plants typically reduces the number of species, although plant sizes can be quite large as one or a few plants come to dominate late-successional mounds. This invasion–abandonment cycle results in a shifting mosaic of mound types through time across the landscape (Fig. 8a). Although the species diversity on mounds changes through time, and the location of mound types varies spatially across the landscape, the total numbers of species and patch types remain constant on the scale of the landscape (Fig. 8c). Therefore, landscape diversity both reflects and determines patterns in diversity at smaller levels of organization, and in particular species diversity.

VIII. CONCLUSIONS

Although much of the current emphasis on biodiversity has been at the level of species, landscape diversity is also important. The preservation and maintenance of multiple levels of organization, including species, populations, communities, and ecosystems, require an understanding of how these various levels interact with their environment across a range of spatial scales. Maintenance of landscape diversity provides a spatial template for the preservation of these smaller levels of organization, and in particular for species biodiversity. Changes in landscape structure and function through time have important effects on the distribution of resources, with resulting influences on the survival of species. Because of the overwhelming numbers of species, it may be impractical to attempt to conserve species diversity per se. By focusing on landscape diversity and the perpetuation of dynamic processes across multiple scales, an attempt can be made to preserve entire ecosystems with their full complement of genetic diversity.

See Also the Following Articles

DEFORESTATION AND LAND CLEARING • DESERTIFICATION • LAND USE ISSUES • KEYSTONE SPECIES • RANGE ECOLOGY • TEMPERATRE GRASSLAND AND SHRUBLAND ECOSYSTEMS

Bibliography

Bierregaard, R. O. J., T. E. Lovejoy, V. Kapos, A. A. dos Santos, and R. W. Hutchings. (1992). The biological dynamics of tropical rainforest fragments. *BioScience* 42, 859–866.

Bormann, F. H., and G. F. Likens. (1979). *Pattern and Process in a Forested Ecosystem: Disturbance, Development and the Steady State Based on the Hubbard Brook Ecosystem.* Springer-Verlag, New York.

Burke, I. C., T. G. F. Kittel, W. K. Lauenroth, P. Snook, C. M. Yonker, and W. J. Parton. (1991). Regional analysis of the central Great Plains. *BioScience* 41(10), 685–692.

Burnett, M. R., P. V. August, J. H. Brown, Jr., and K. T. Killingbeck. (1998). The influence of geomorphological heterogeneity on biodiversity. I. Patch-scale perspective. *Conservation Biol.* 12(2), 363–370.

Coffin, D. P. (1997). Spatially-explicit propagule sources and vegetation dynamics at a grassland–shrubland ecotone. US–International Association for Landscape Ecology conference, March 16–20, 1997, Duke University, Durham, North Carolina.

Coffin, D. P., and W. K. Lauenroth. (1994). Successional dynamics of a semiarid grassland: Effects of soil texture and disturbance size. *Vegetatio* 110, 67–82.

Coughenour, M. B. (1991). Spatial components of plant–herbivore interactions in pastoral, ranching, and native ungulate ecosystems. *J. Range Management* 44(6), 530–542.

Forman, R. T. T. (1995). *Landscape Mosaics: The Ecology of Landscapes and Regions.* Cambridge University Press, Cambridge, United Kingdom.

Forman, R. T. T., and M. Godron. (1986). *Landscape Ecology.* John Wiley & Sons, New York.

Fuller, J. L., D. R. Foster, J. S. McLachlan, and N. Drake. (1998). Impact of human activity on regional forest composition and dynamics in central New England. *Ecosystems* 1, 76–95.

Gosz, J. R. (1995). Edges and natural resource management: Future directions. *Ecology Int.* 22, 17–34.

Huston, M. A. (1994). *Biological Diversity: The Coexistence of Species on Changing Landscapes.* Cambridge University Press, Cambridge, United Kingdom.

Knight, D. H., and L. L. Wallace. (1989). The Yellowstone fires: Issues in landscape ecology. *BioScience* 39, 700–706.

Kroel-Dulay, G., T. Hochstrasser, and D. P. Coffin. (1997). Compositional comparison of grass-dominated patches at a semiarid–arid ecotone. *Bull. Ecol. Soc. Amer.* 78, 272.

Lapin, M., and B. V. Barnes. (1995). Using the landscape ecosystem approach to assess species and ecosystem diversity. *Conservation Biol.* 9(5), 1148–1158.

Lovejoy, T. E., J. M. Rankin, R. O. Bierregaard, Jr., K. S. Brown, Jr., L. H. Emmons, and M. Van der Voort. (1984). Ecosystem decay of Amazon forest remnants. In *Extinctions* (M. H. Niteki, ed.), pp. 295–325. Westview Press, Boulder, Colorado.

Lovejoy, T. E., R. O. Bierregaard, Jr., A. B. Rylands, J. R. Malcolm, C. E. Quintela, L. H. Harper, K. S. Brown, Jr., A. H. Powell, G. V. H. Powell, H. O. R. Schubart, and M. B. Hays. (1986). Edge and other effects of isolation on Amazon forest fragments. In *Conservation Biology: The Science of Scarcity and Diversity* (M. Soule, ed.), pp. 251–285. Sinauer Associates, Sunderland, Massachusetts.

Mack, R. N. (1981). Invasion of *Bromus tectorum* L. into western North America: An ecological chronicle. *Agro-Ecosystems* 7, 145–165.

Milne, B. T. (1988). Measuring the fractal dimension of landscapes. *Appl. Math. Computations* 27, 67–79.

Neilson, R. P., and R. J. Drapek. (1998). Potentially complex biosphere responses to transient global warming. *Global Change Biol.* 4, 505–521.

Nichols, W. F., K. T. Killingbeck, and P. V. August. (1998). The influence of geomorphological hetrogeneity on biodiversity. II. A landscape perspective. *Conservation Biol.* 12(2), 371–379.

O'Neill, R. V., J. R. Krummel, R. H. Gardner, G. Sugihara, B. Jackson, D. L. DeAngelis, B. T. Milne, M. G. Turner, B. Zygmunt, S. W. Christensen, V. H. Dale, and R. L. Graham. (1988). Indices of landscape pattern. *Landscape Ecol.* 1, 153–162.

Pastor, J., and W. M. Post (1988). Response of northern forests to CO_2-induced climate change. *Nature* 344, 55–58.

Pickett, S. T. A., and J. N. Thompson. (1978). Patch dynamics and the size of nature reserves. *Bio. Conservation* 13, 27–37.

Pickett, S. T. A., and P. S. White (eds.). (1985). *The Ecology of Natural Disturbance and Patch Dynamics.* Academic Press, New York.

Platt, W. J. (1975). The colonization and formation of equilibrium plant species associations on badger disturbances in a tall-grass prairie. *Ecol. Monographs* 45, 285–305.

Prentice, C., W. Cramer, S. P. Harrison, R. Leemans, R. A. Monserud, and A. M. Solomon. (1992). A global biome model based on plant physiology and dominance, soil properties and climate. *J. Biogeogr.* 19, 117–134.

Riitters, K. H., R. V. O'Neill, C. T. Hunsaker, J. D. Wickham, D. H. Yankee, S. P. Timmins, K. B. Jones, and B. L. Jackson. (1995). A factor analysis of landscape pattern and structure metrics. *Landscape Ecol.* 10, 23–40.

Romme, W. H. (1982). Fire and landscape diversity in subalpine forests of Yellowstone National Park. *Ecol. Monographs.* 52, 199–221.

Romme, W. H., and D. H. Knight. (1982). Landscape diversity: The concept applied to Yellowstone Park. *BioScience* 32, 664–670.

Rowe, J. S. (1992). The ecosystem approach to forestland management. *Forest Chronicle* 68, 222–224.

Saunders, D. A., and R. J. Hobbs (eds.). (1991). *The Role of Corridors.* Surrey Beaty & Sons, Chipping Norton, Australia.

Saunders, D. A., R. J. Hobbs, and C. R. Margules. (1991). Biological consequences of ecosystem fragmentation: A review. *Conservation Biol.* 5, 18–32.

Schulze, E. D., and H. A. Mooney (eds.). (1994). *Biodiversity and Ecosystem Function.* Springer-Verlag, Heidelberg, Germany.

Turner, M. G. (1989). Landscape ecology: The effect of pattern on process. *Annu. Rev. Ecol. Systematics* 20, 171–197.

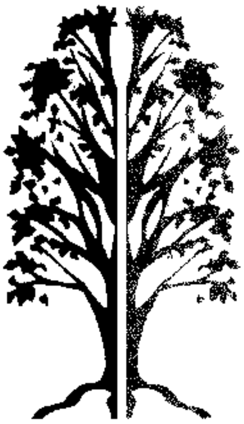
Turner, M. G., and R. H. Gardner. (1990). *Quantitative Methods in Landscape Ecology. The Analysis and Interpretation of Landscape Heterogeneity.* Springer-Verlag, New York.

Turner, M. G., V. H. Dale, and R. H. Gardner. (1989). Predicting across scales: Theory development and testing. *Landscape Ecol.* 3, 245–252.

Whittaker, R. H. (1960). Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monographs* 30, 279–338.

Whittaker, R. H. (1972). Evolution and measurement of species diversity. *Taxon* 21, 213–251.

Wickham, J. D., T. G. Wade, K. B. Jones, K. H. Riitters, and R. V. O'Neill. (1995). Diversity of ecological communities of the United States. *Vegetatio* 119, 91–100.



LAND USE ISSUES

John M. Marzluff and Nathalie Hamel
University of Washington

GLOSSARY

afforestation Process of planting trees in areas currently devoid of them. It is done in desert regions and is increasingly common in Europe, where grasslands and pastures are being converted into woodlands and forests.

agricultural intensification Management of agricultural land to increase yield principally through the use of high-yielding crop varieties, chemical fertilizers and pesticides, irrigation, shorter rotations, larger fields, and mechanization.

desertification Reduction of the biological potential of land that ultimately leads to the creation of desertlike conditions. It occurs primarily through overuse of marginal lands by humans for agriculture.

grassland Land that has a vegetation cover dominated by grasses.

habitat fragmentation Splintering of once contiguous land cover into isolated pieces. Fragmentation happens when habitat is lost from the interior, rather than the edge, of a large block of cover. The resulting habitat patches are sometimes referred to as "habitat islands" and the intervening, converted land is called the "matrix."

land cover Physical and biotic character of the earth's terrestrial surface. Land cover is typically the vegetation (e.g., tropical forest, marsh, desert, cornfield, shrubland) or human construct (e.g., road, dwelling, industrial area) that covers the surface. It may be grossly defined as in the preceding examples, or defined at finer scales (e.g., moist deciduous tropical forest, moist evergreen tropical forest, dry tropical forest).

land use The use that humans make of the earth's terrestrial surface, usually to obtain goods or benefits. Examples include agriculture, mineral exploration, settlement, natural reserve, and timber production. Land use refers to the dominant human *activity* that occurs on the earth's surface, and it often modifies natural characters to change land cover (e.g., replacing native grassland with a cornfield or a road).

settlement Land used for human habitation, including for the construction of cities, towns, villages, rural settlements, and roads. In a gross sense, settlement is also a land cover when used to refer to the land occupied by the variety of trappings accompanying human living space, including homes, gardens, roads, transportation centers, and industrial areas.

LAND USE ISSUES CONCERN THE PROCESSES by which human activities determine land cover. Important issues are agricultural development and intensification, settlement, and extraction of natural resources.

In response to human land use, the earth's land cover has changed from a mosaic of native woodlands, forests, and grasslands to an increasingly impacted mixture of degraded and fragmented native habitats, exotic croplands, and impervious urban surfaces. In the last three centuries, models suggest that forests have declined 19%, grasslands have declined 8%, and cropland has increased over 400%. This article discusses how land use processes have changed through time and how they have caused the natural pattern of land cover to change. This transformation of the planet's landscape is widely recognized as the primary driver in the current global loss of biodiversity. Several examples of how land use can influence biodiversity are also considered.

I. LAND USE ISSUES

Humans obtain food, energy, and shelter from the earth by harvesting natural resources and creating new resources. Both of these activities, especially the creation of new resources by agricultural and industrial actions, modify our environment at a scale unknown for other organisms. Our unparalleled recent success as a species (increasing in population size by nearly 2% per year from 1950 to 1985) and our tremendous ability to modify the environment have allowed us to dramatically transform the earth (Turner *et al.*, 1990). It is estimated that we have transformed or degraded from 39 to 50% of the earth's surface and now command 25% of the world's total net primary productivity (Vitousek *et al.*, 1986, 1997).

We can conceptualize our effects on the earth's surface by relating our land and resource use to changes in land cover (Fig. 1). As our population increases, we garner more of the earth to supply our needs. The aspects of the human population that drive changes in land use and resulting land cover are known only in a general way. Population size is obviously an important determinant of our land use. But its importance relative to global affluence, technological capacity, political organization, and social structuring is hotly contested (Meyer and Turner, 1992). Recent insights suggest that population size may directly determine land use at a global scale, but land use locally is determined by the interaction of regional population size, technology, socioeconomic development, industrialization, and policy (see Fig. 1). Total biological demand for resources and resulting conversion of land cover are determined by the total world population. Where on the planet changes in land use occur is affected by regional population size, technology, policies, conventions, cultural attitudes, and economic forces that discourage or reward land conversion at the local scale (Meyer and Turner, 1992).

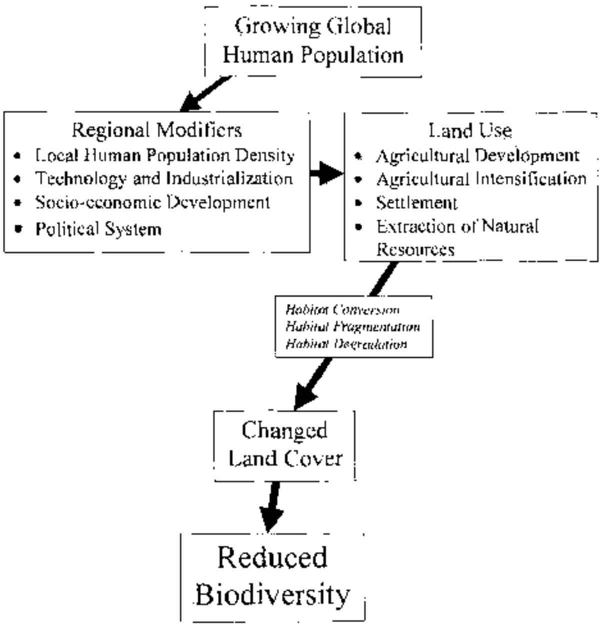


FIGURE 1 The growing global population of humans reduces biodiversity primarily through land use that converts, fragments, and degrades land cover. Human population growth directly affects land use by demanding more food, living space, and natural resources. The specific effects on regional land use are modified by the local human density, local economy, local technology and industrialization, and local political system.

Three main land uses are responsible for changes in the earth's land cover. These are (1) *agriculture*, which includes cultivation, grazing, irrigation, and drainage; (2) *extraction of natural resources*, primarily for forestry and energy production; and (3) *settlement*, which we define loosely to include use of land for human settlements ranging from rural villages to cities. Our domination of the earth is clearly reflected by the earth's current land cover, which is primarily a mixture of fragmented and degraded native habitats, exotic croplands and pastures, and settlements (Fig. 2). This transformation of land cover per se is recognized as the driving force in the loss of biological diversity worldwide. Additionally, land use affects the biosphere indirectly by modifying ecosystem processes and introducing exotic species (Vitousek *et al.*, 1997).

A. Agriculture

Our agricultural practices have had the greatest influence on the earth's land cover. Before the development of agricultural society, several lines of evidence suggest

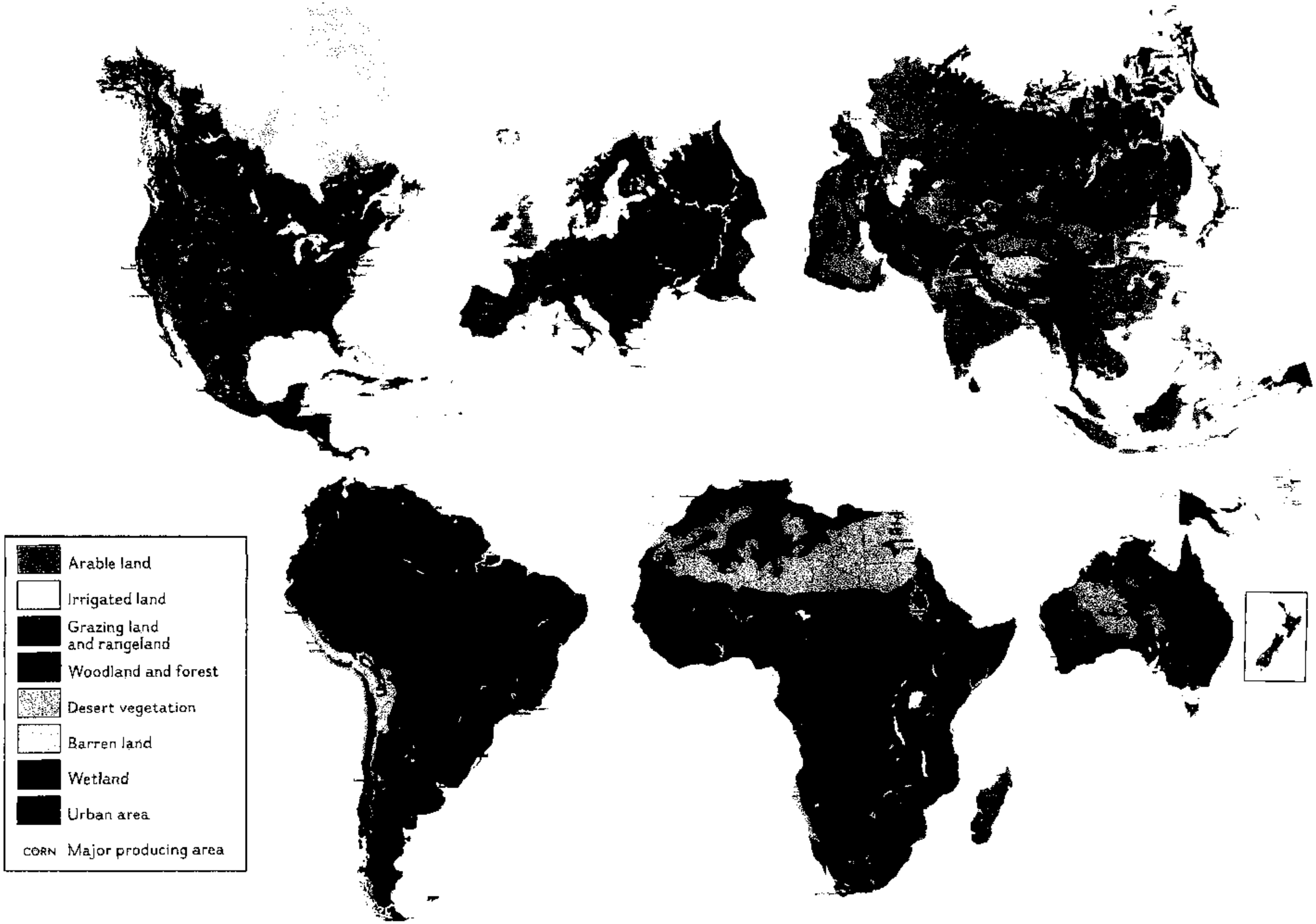


FIGURE 2 Modern land cover (gross scale) of the world. Human domination is evident by the preponderance of agricultural land cover (arable, irrigated, grazing, and range lands). (From *The National Geographic Atlas of the World*, National Geographic Society, Washington, D.C., 1992.)

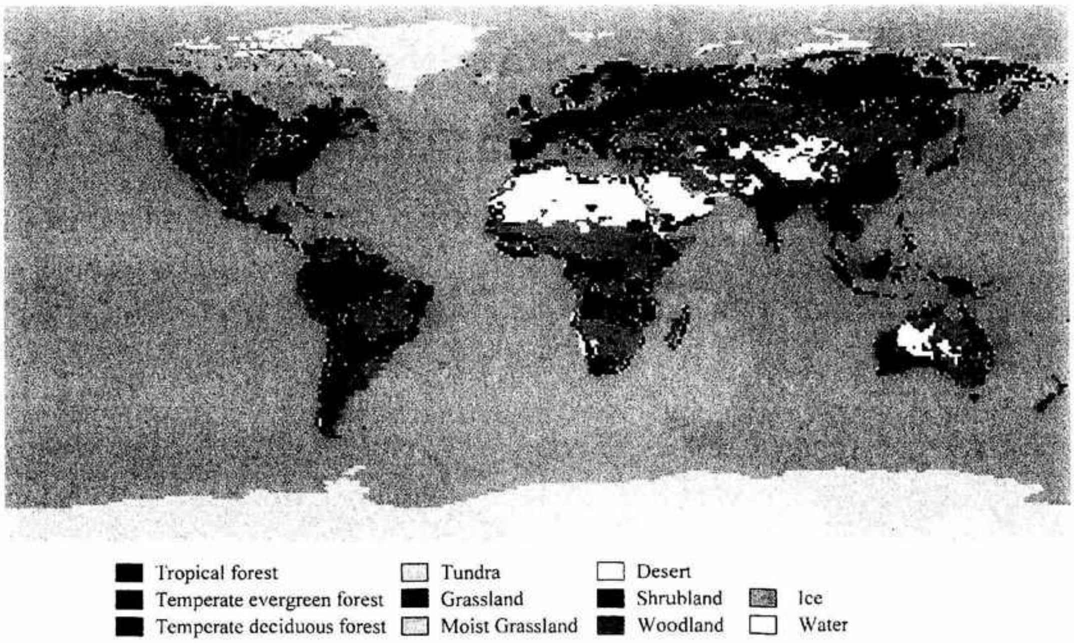


FIGURE 3 Estimated land cover before the development of agriculture. Climatic conditions were modeled by Matthews (1983) to estimate land cover. We created this image by aggregating the major UNESCO vegetation classes. (From E. Matthews, *Global Vegetation*, NASA Goddard Institute for Space Studies, Washington, D.C., 1983.)

that the earth was covered with a complex mosaic of forest, woodland, grassland, wetland, tundra, and desert (Fig. 3). During the last 300 years we have cultivated approximately 12% of the earth's surface. This represents an approximate increase of 390 to 466% since 1700 (Fig. 4; Richards, 1990). Increases in cultivated land came mostly at the expense of native forests (other than tropical rain forest) and native grasslands (Fig. 5). We transform the earth by burning and clearing forests, tilling grasslands, draining wetlands, and irrigating arid lands. Agricultural development is responsible for the overwhelming majority of wetland conversion (the

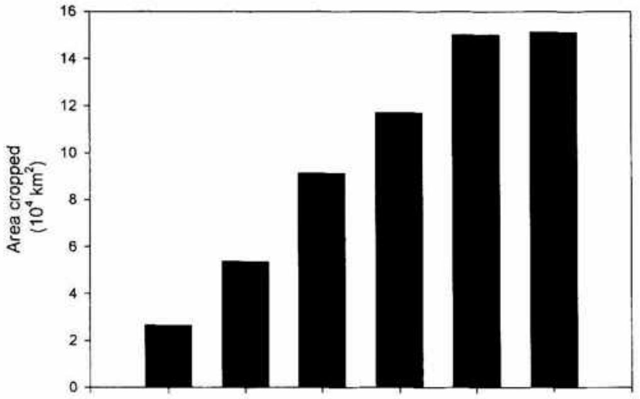


FIGURE 4 Amount of the world under cultivation (cropland) from 1700 to 1996. (From Richards, 1990, updated by 1996 FAO statistics.)

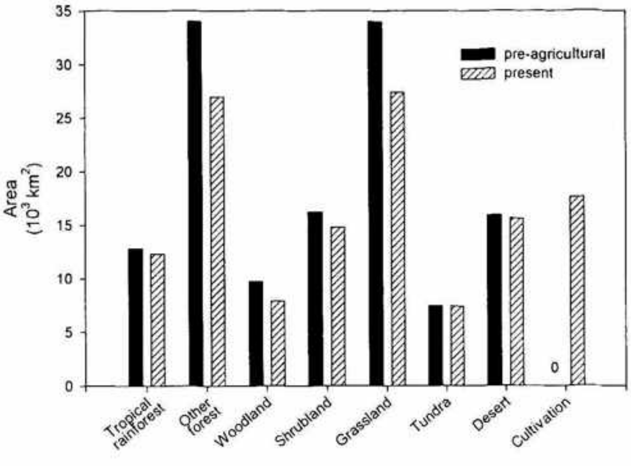


FIGURE 5 Areas occupied by major ecosystems of the world before agriculture (Matthews, 1983) and in 1991 (Meyer and Turner, 1994). Cultivation has mostly affected temperate and tropical forests other than rain forests, as well as grassland ecosystems.

global loss of wetlands is not accurately estimated; Meyer and Turner, 1992). The pattern of increasing human population, increasing cropland, and declining forests and grasslands is consistent across the continents except in Europe, where current afforestation is increasing forest and woodland cover (Fig. 6).

Although conversion of native land to exotic croplands has slowed, our use of agricultural land has intensified since "the Green Revolution" began in the 1960s (Matson *et al.*, 1997). Agricultural intensification alters land cover by (1) creating highly productive monocultures, (2) diverting water resources, and (3) fueling technology to produce better seeds and more fertilizers, pesticides, and insecticides.

Increasingly intensive use of grasslands for agriculture has resulted in worldwide overgrazing (Fig. 7). A direct effect of this is "desertification" or the degradation of land to desertlike conditions. Current estimates indicate that 6% of the earth's land has been converted to deserts by agricultural activity; nearly 25% of the earth is threatened with desertification (Meyer and Turner, 1992).

B. Natural Resource Extraction

We extract a variety of resources from the earth for shelter and energy, but at a global scale timber harvest and management have the largest effect among extractive activities on landscape composition and land cover. Roughly one-third of the earth is forested and 18% (650 million ha) of forests are currently maintained as plantations or managed for goods and services (Noble and Dirzo, 1997). Humans affect all forests by hunting, gathering, recreating, and suppressing fire. However, only fire suppression has the potential to change forest cover type (if successful, it initially slows forest growth by reducing natural thinning but may allow for greater development of late-successional forests by reducing disturbance).

Natural resource extraction has a less direct effect on forest cover than does agriculture. Resource extraction modifies the forest by changing seral stage, degrading forest integrity, and fragmenting continuous forest, but is associated with reduced forest cover in only a few locations (southeastern Asia and parts of Africa, Latin America, and India). Agriculture, in contrast, is the primary agent of forest loss, although it also modifies forest function. Some of the forest loss to agriculture is compensated for by maintenance and planting of forest for timber. From 1980 to 1995, one-tenth of the 13 millon ha of forest lost per year to agriculture was compensated for by the creation of forest plantations. Much of this trade-off is occurring in Europe, where afforestation campaigns are increasing forest land at the expense of cropland (see Fig. 6c).

C. Settlement

The clearing of native land cover and drainage of wetlands for human settlement are the most permanent forms of land conversion. Overall, 2.5 to 6% of the earth is now occupied by human dwellings and their supporting infrastructure (roads, industrial areas, reservoirs, etc.). About 10% of this area is actually built area with impervious surface (Meyer and Turner, 1992). Human settlements typically occupy less than 20% of even the most developed country's area, but this varies widely (Fig. 8). Each night we advertise our settlements by turning on lights. Viewing this spectacle from the air provides a striking representation of the degree to which settlements occupy the earth (Fig. 9). Human settlement affects nearly all land in Europe, eastern North America, India, and parts of Asia.

Settlement has increased dramatically over the last 300 years as evidenced by the increasing number of large cities (Fig. 10). As human populations have grown, we have settled with increasing frequency in urban areas. Today nearly half of all humans live in urban areas (primarily in the developed countries), compared to only 14% at the start of the twentieth century (Meyer and Turner, 1992). A result of our increasingly urban life is an increased occupation of area by cities and surrounding lower-density housing and industrial areas (urban and suburban sprawl). This expansion of urban areas in developed countries and the increasing density of large urban areas in the developing tropics are responsible for the dramatic and likely irreversible transformation of the earth's surface.

D. History Repeats Itself: Past and Future Effects of Land Use on Land Cover

Our transformation of the earth during the last five centuries has proceeded by pioneers expanding frontier settlements in a fairly regular way. When humans settle an area, they open a new frontier where natural resources are abundant, cheap, and readily obtained. As resources are consumed and population increases, we expand our "footprint" on the surrounding land in a series of radiating waves. We transform the land at successively greater distances from a settlement's center by clearing and draining land, extracting resources, and converting land cover to suit our needs (Berry, 1990).

The conversion of land cover was fueled on a global

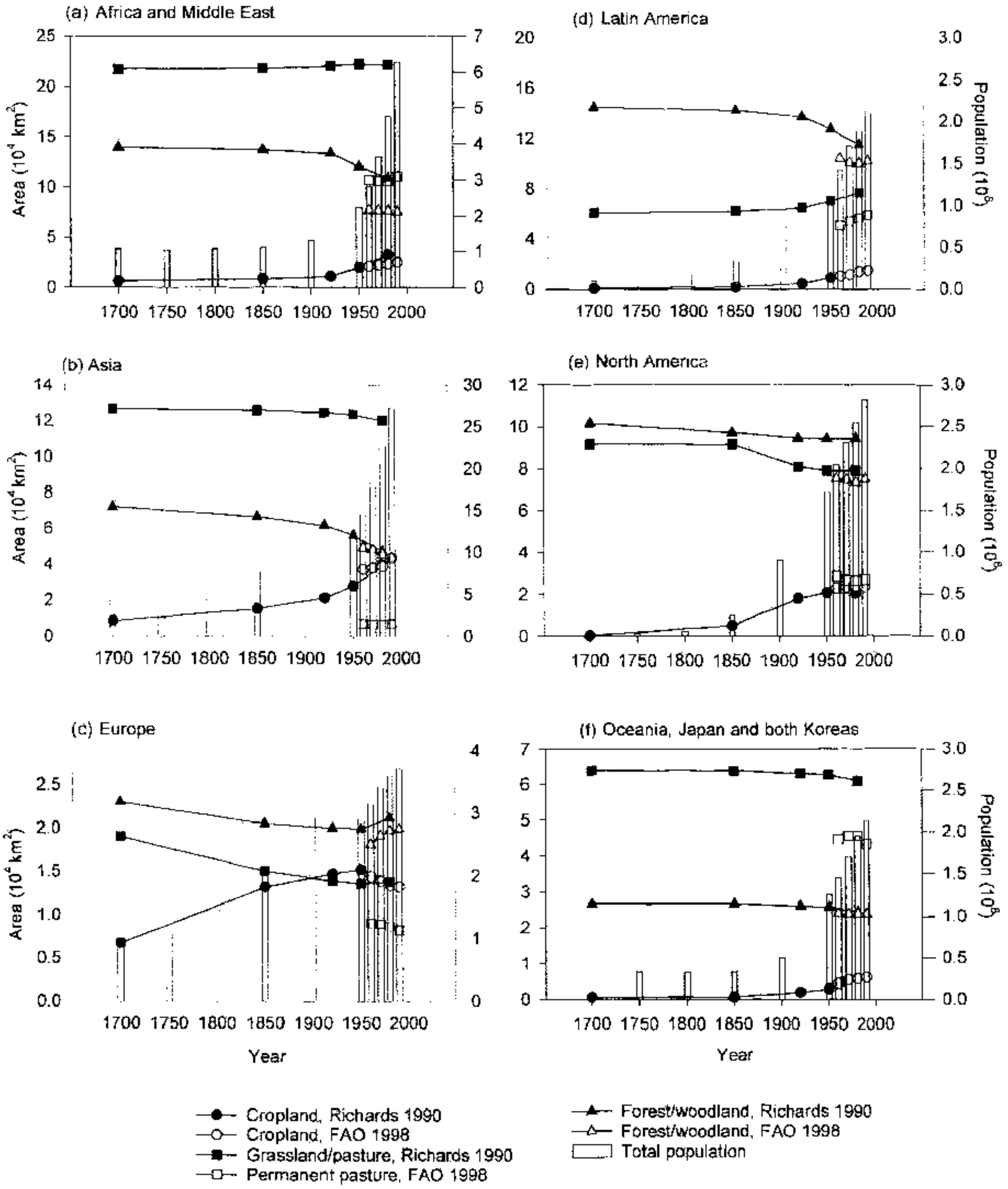


FIGURE 6 Land use trends and population estimates of major regions of the earth. Land use trends between 1700 and 1980 (closed symbols) are compiled from Richards (1990). After 1950 (open symbols), land use estimates are derived from FAOSTAT (www.fao.org, October 1998). Population estimates in (a) are for Africa only. Panel (f) includes Japan, North and South Korea, and Oceania.

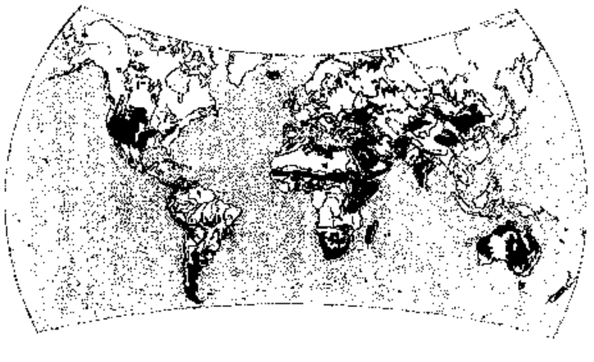


FIGURE 7 Areas of the world affected by overgrazing. This human-induced activity has been identified as one of the major causes for soil degradation. Overgrazed areas are the result of actual removal of biomass by grazing animals and the effect of trampling. The effects are enhanced where too many animals or the wrong kind of animals are grazed. (From UNEP/ISRIC, *World Atlas of Desertification*, p. 44, United Nations Environment Programme, published by Arnold/John Wiley & Sons, New York, 1997.)

scale by Western European nations and their seaborne pioneers from the 1700s to the early 1900s (Richards, 1990). Once the European frontiers were settled, the appetite for amenity and subsistence resources propelled the opening of new frontiers around the world. These frontiers then had to satisfy local and European needs, so resources were harvested with frenzied zeal as opportunities seemed boundless on the frontier. Landscapes were quickly degraded and transformed, and as resources dwindled, the frontier closed and pioneers moved on to new frontiers. However, the old frontier was rarely left to return to a natural state. Rather, a more orderly settlement usually persisted, leading to slower, but more permanent, land transformation. Many of these settlements became new centers in a changing world economy, thereby broadening the base from which remaining frontiers were exploited.

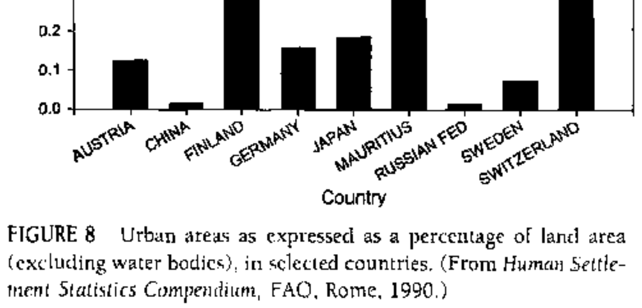


FIGURE 8 Urban areas as expressed as a percentage of land area (excluding water bodies), in selected countries. (From *Human Settlement Statistics Compendium*, FAO, Rome, 1990.)

The global pattern of urbanization over the last 300 years (see Fig. 10) illustrates the consequences of our opening and closing the earth's frontiers. New frontiers open in areas increasingly far from Europe, and develop large urban areas as they close. The frontiers of Europe, eastern North America, India, and Asia have closed, leaving dense human settlement to permanently change land cover. As the remote frontiers of the tropics, mountainous regions, and northern forests and tundras are exploited, settlement will likely increase and land will be transformed to a lasting urban state. Europe can be viewed as a crystal ball for much of the earth's future land use. The forecast is for drastic simplification of a once diverse native land cover to one dominated by settlements, intensive agriculture, and fragmented vestiges of native habitat.

II. EFFECTS OF LAND USE ON BIODIVERSITY

The current extinction crisis is fueled, at a proximate level, by (1) habitat conversion, (2) habitat fragmentation, and (3) habitat degradation. These are the most common reasons given to explain why modern species are at risk of extinction (IUCN, 1996). These three processes are the direct result of our land use and they function as the mechanisms that connect changing land cover to biodiversity (see Fig. 1). Conversion of native land cover to agriculture and settlement is unquestionably the main cause of habitat loss on a global scale. This conversion typically occurs in a piecemeal fashion and is associated with extensive road building that fragments habitat as it is being lost. Native and converted habitats that are already utilized by humans are further degraded by agricultural intensification, resource extraction, and recreation.

The effects of land use on biodiversity can be illustrated by considering habitat conversion, fragmentation, and degradation in several regional studies, and we have selected well-documented studies to illustrate the general process. It is important to consider regional changes in land cover because the factors that drive land use and link it to land cover are best understood

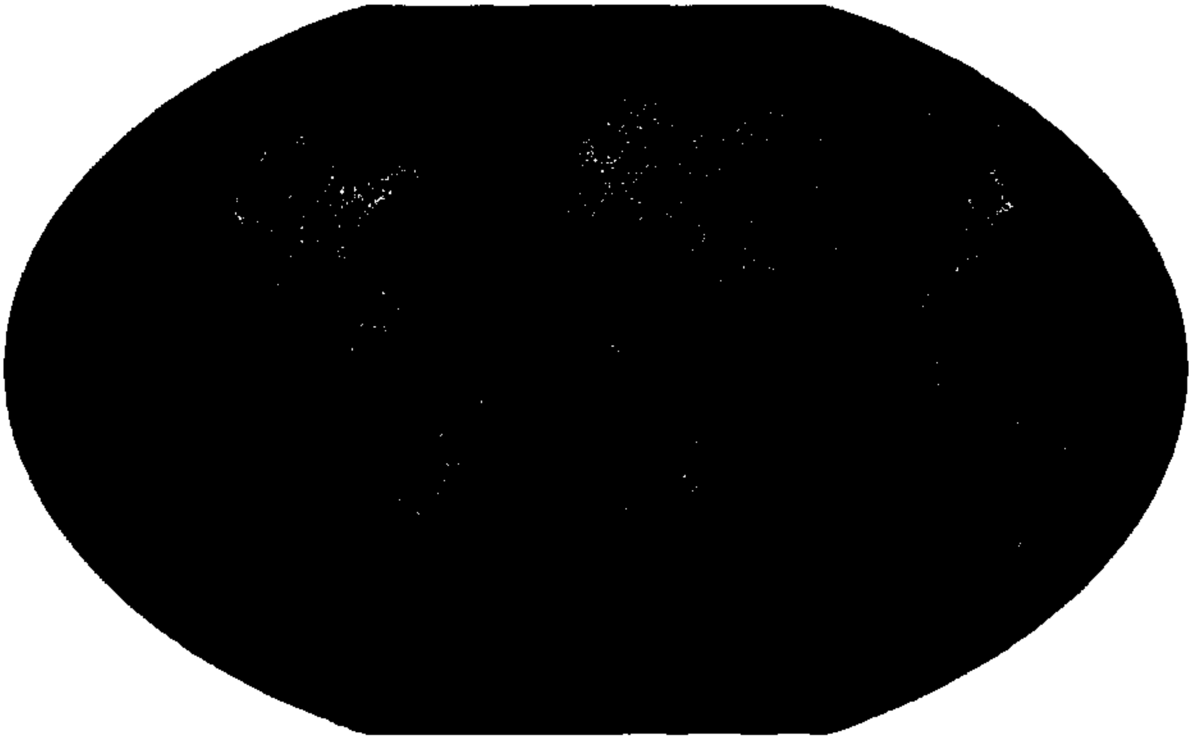


FIGURE 9 The world at night. Viewed from space, the distribution of lights (white blurs) in human settlements indicates the omnipresence of urban space. This image taken from a satellite is a composite of cloud-free observations made over a 6-month period. The red is the result of fires, and the yellow indicates natural gas flares. (From *National Geographic Maps*, October 1998, National Geographic Society, Washington, D.C.)

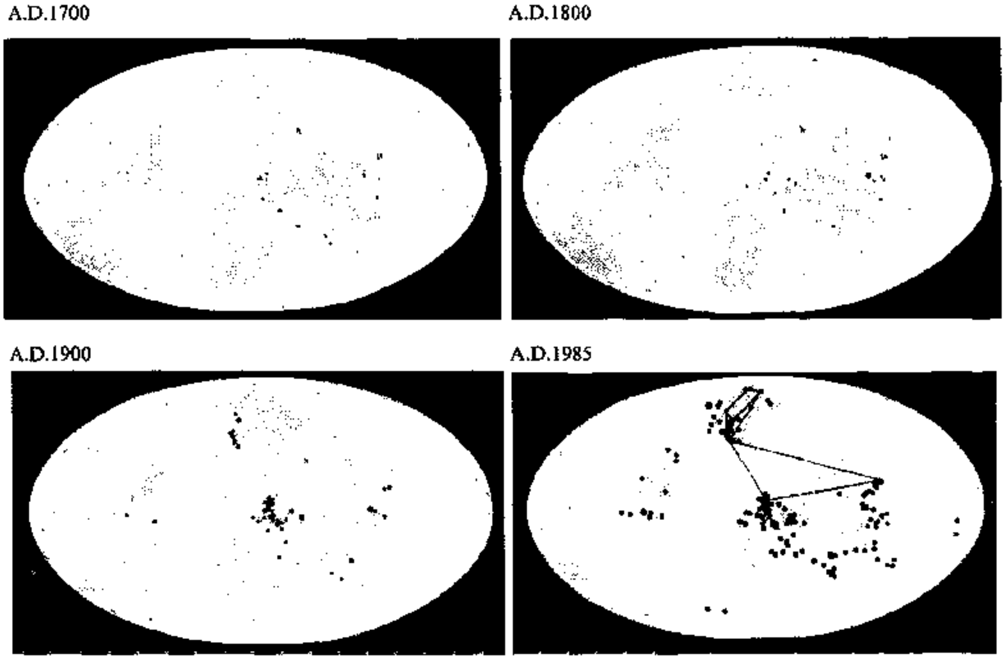


FIGURE 10 Distribution of the sizes of world cities, between 1700 and 1985. The growth rate of cities in the Americas is more rapid relative to that of cities on the European and East Asian continents. The black dots represent cities of varying sizes (●, 200,000 to 500,000; ●, 500,000 to 1,000,000; ●, 1,000,000 to 10,000,000). From Berry (1990).

at "subglobal" levels (Kummer and Turner, 1994). The following sections examine (1) the complex ways that biophysical and socioeconomic factors affect land use, (2) the linkage between land use and habitat conversion, fragmentation, and degradation, and (3) the effects of land use on biodiversity.

A. Habitat Conversion

1. Deforestation in Southeast Asia

The tropical rain forests of Southeast Asia have declined dramatically since the end of World War II (in the late 1940s). Before humans arrived, the Philippines, Indonesia, Malaysia, Vietnam, and Thailand were completely forested (see Fig. 3). Their rich dipterocarp forests were heavily logged (primarily for export) after the war ended. The Philippines currently retains less than 20% of its original forest cover; Malaysia retains more forest, but it is being extensively harvested (Fig. 11). The rates of deforestation in the Philippines are well documented and staggering (Collins *et al.*, 1991; Kummer, 1992). From 1948 to 1987, 55% of the forest was lost. From 1969 to 1988, the rate of deforestation was over 2% per year—or 2 ha every 5 minutes!

The process of habitat loss (deforestation) in the Philippines serves as a model for all of Southeast Asia (Kummer and Turner, 1994). Two linked land uses, logging and subsistence agriculture, are responsible for converting land cover from tropical rain forest to arable land. Logging dipterocarps converted the primary rain forest to a simpler secondary forest. The resulting partial clearing and road construction allowed poor, landless people to remove the secondary forest and expand agriculture.

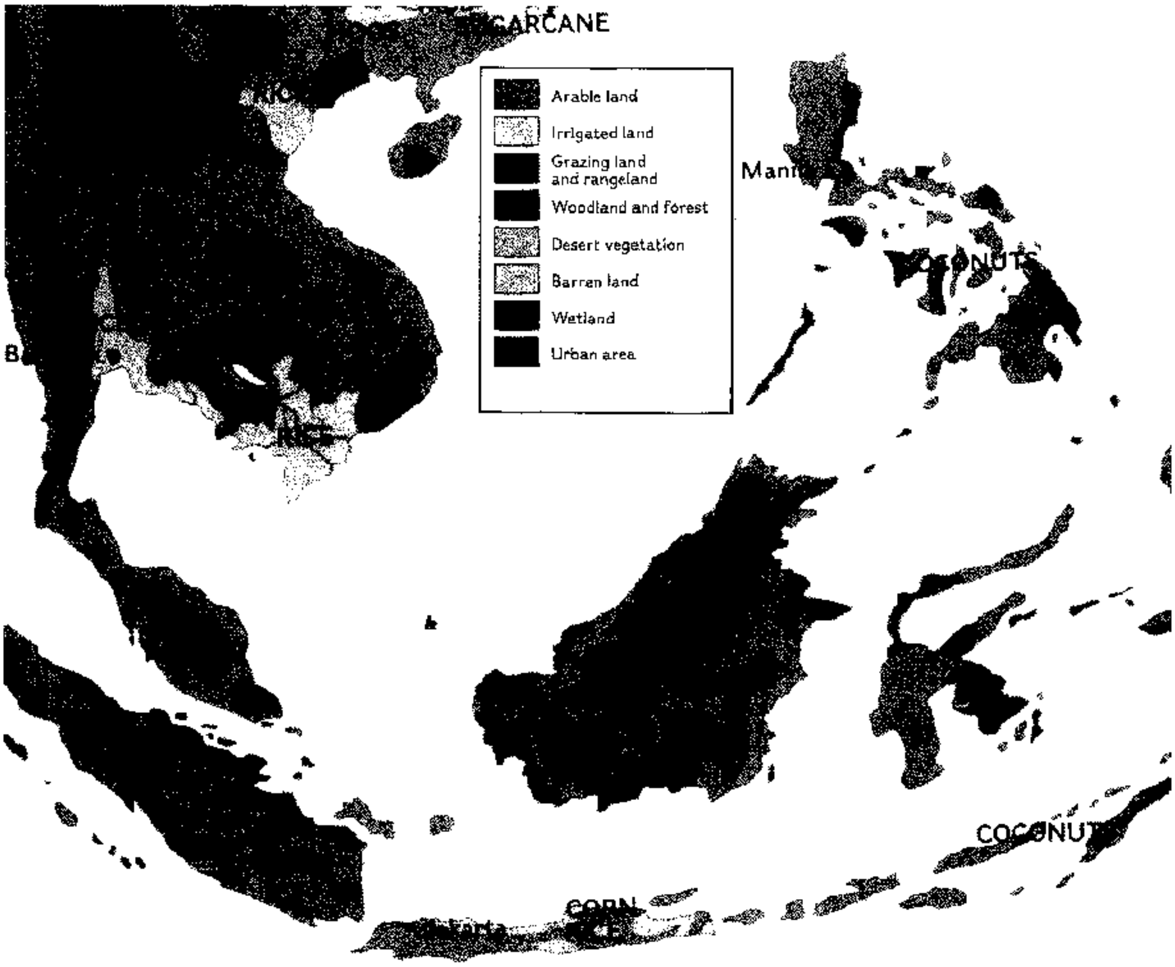


FIGURE 11 Current land cover in Southeast Asia. The tropical rain forests that once covered the area are now reduced to patches in an agricultural landscape. (From *The National Geographic Atlas of the World*. National Geographic Society, Washington, D.C., 1992.)

This two-step process illustrates how human density, global markets, and local socioeconomic conditions work together to cause changes in land cover. Despite the Philippines' extremely dense (217 people/km^2 in 1993) and rapidly growing population (2.5% in 1993), deforestation is not directly driven by the local population. Rather, a global demand for wood, coupled with corrupt and unrestricted logging, cleared land and made it accessible to people in need of agricultural space. A surplus of landless people was needed to drive the forest conversion to agricultural land, but without logging to open a new frontier, urbanization might have increased and agriculture might have remained more centralized and intense rather than dispersed as it is now.

Southeast Asia in general, and the Philippines in particular, has tremendous biodiversity. Diversity in the Philippines is poorly known, but there are >12,000 plants and fungi and >950 terrestrial vertebrates. This incredible diversity is primarily found in a large number (>59) of scattered and poorly protected national parks. Less than 1.3% of the Philippines is managed for conservation, and much of this area is still logged and settled by landless agriculturalists.

The combination of high biodiversity, high endemism, and extreme habitat conversion results in extreme risks of extinction. The magnitude of threat to birds and mammals in Southeast Asia is particularly well known (IUCN, 1996). Worldwide, Indonesia has the largest number of birds (104) and mammals (128) threatened with extinction. The Philippines leads the world in the percentage of its avifauna that is threatened with extinction (15%). These countries are second only to Madagascar in the percentage of mammals threatened with extinction (32%). Vietnam, Thailand, and Malaysia are all in the world's top 20 countries for species endangerment.

To summarize, deforestation in Southeast Asia is an example of how our land use for amenity (dipterocarp logging) and subsistence (agriculture) resources changes land cover and affects biodiversity. Habitat conversion threatens biodiversity, but the greed (exotic wood) and need (agriculture) of a growing global human population fuel the conversion.

2. Urbanization in Western Washington State

An expanding population and the world's appetite for resources have transformed much of North America from extensive forest and grassland to agricultural and urban land (compare Figs. 2 and 3). The general story is similar to the case in Southeast Asia. However, in the Pacific Northwest a more dramatic and permanent conversion of land is occurring at a rapid rate. Logging during the last 50 years has converted more than two-thirds of the primary forest in Washington to secondary forest. Rather than continuing to manage these forests for timber production, much of the land is now being converted to human settlements. In western Washington, for example, human populations have doubled in the last 50 years and are expected to double again in the next 50 years. From 1998 to 2000, Washington is expecting a net gain of one person every 5 minutes! This places a premium value on land for settlement. The predictable result is a rapid conversion of forest to urban and suburban settlements. Indeed, from 1970 to 1997, Washington lost 2.3 million acres of managed forestland. Urban expansion was responsible for about half of this loss. Rights-of-way and agriculture accounted for the rest.

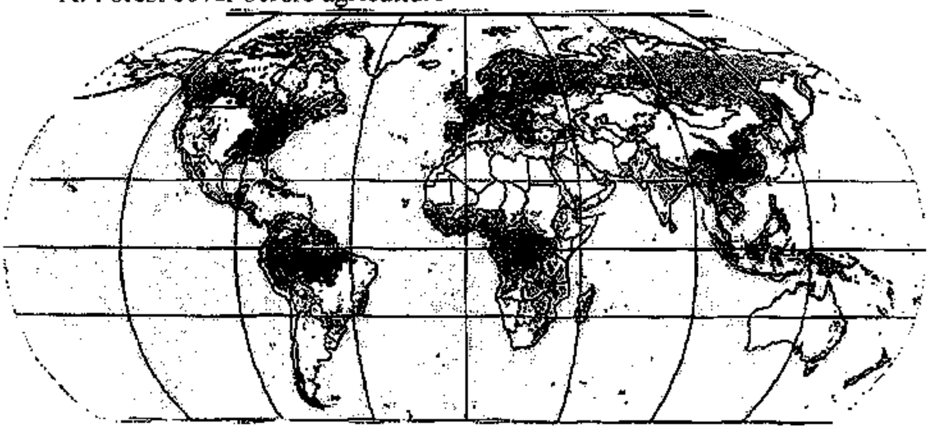
The increasing urbanization in western Washington threatens to reduce biodiversity. Water flows altered by settlement have reduced the spawning and rearing habitat for the region's spectacular salmon diversity. As a result, several salmon runs have been extinguished and many are now listed as endangered. Loss of forests and intensification of resource extraction and recreation on remaining forests have contributed to the endangerment of several birds (e.g., Spotted Owls, *Strix occidentalis*, and Marbled Murrelets, *Brachyramphus marmoratus*), mammals (e.g., Grizzly Bear, *Ursus arctos*, and Gray Wolf, *Canis lupus*), and amphibians (Larch Mountain Salamander, *Plethodon larselli*). Drainage of wetlands and settlement of native woodlands and grasslands have endangered amphibians (Oregon Spotted Frog, *Rana luteiventris*), insects (Oregon Silverspot butterfly, *Speyeria zerene*), mammals (Western Gray Squirrel, *Sciurus griseus*), and reptiles (Western Pond Turtle, *Clemmys marmorata*).

Some species benefit whenever land cover changes. A good example of such a species is the American Crow (*Corvus brachyrhynchos*) in western Washington. Crows are found only in close association with human settlement. As a result, their numbers have increased 10-fold from 1960 to 1995. Increases in human commensals like crows may have a ripple effect through the native biota. Crows prey on the eggs and young of other birds, which may limit their reproduction and reduce overall biodiversity close to human settlements even in remaining forest reserves.

B. Habitat Fragmentation

The loss of native habitat as lands are converted for human use splinters the remaining habitat into small

A. Forest cover before agriculture



B. Modern forest cover

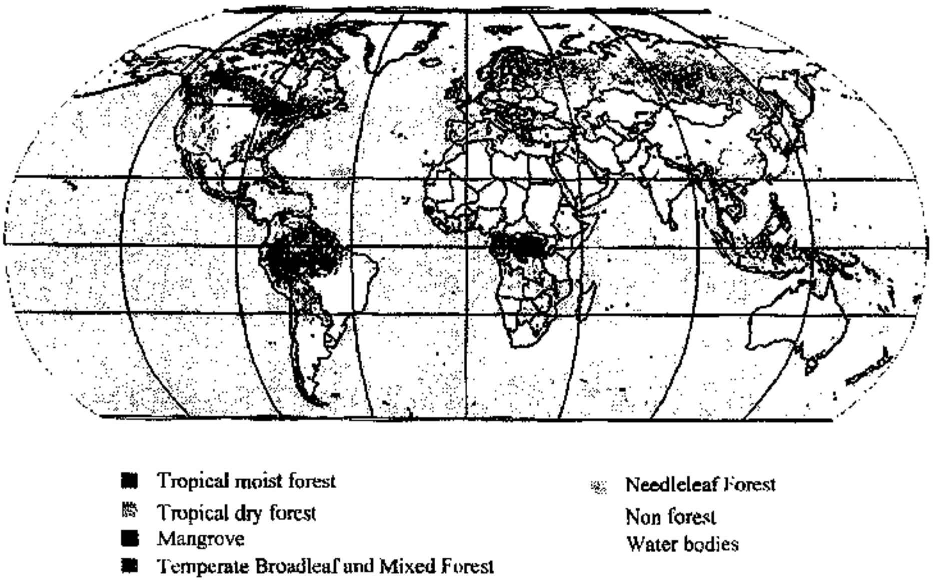


FIGURE 12 Fragmentation of the world's forests at present (B) compared to in preagricultural time (A). The original distribution of forests is estimated from climatic models. Current distributions are based on a variety of sources compiled from the 1980s to early 1990s. (From The World Conservation Monitoring Center, Cambridge, United Kingdom, www.wcmc.org.uk.) See also color insert, Volume 1.

fragments afloat in a human-dominated matrix. Such "habitat fragmentation" is most evident on a global scale when we compare the original and current distributions of forests (Fig. 12). Contiguous primary forests have been removed to varying degrees, resulting in a patchwork of forest "islands" in a "sea" of settlement, agriculture, and second growth forests.

Forest fragmentation affects biodiversity by (1) outright loss of habitat (recall Southeast Asia), (2) creation of forest edges that differ from interior forest in many physical and biological properties (e.g., wind speed, humidity, temperature, and predator populations), and (3) disruption of movement and dispersal patterns of forest species. The effects of habitat fragmentation depend on the type of land cover that surrounds the fragments (the "matrix"). In general, when the matrix is similar to the native habitat (e.g., secondary forest surrounding native forest), the effects of fragmentation are less than when the matrix is dominated by human settlements or agriculture. The size of the remaining fragment is also important; large fragments conserve more biodiversity than small ones.

A variety of recent studies of fragmentation in the tropics suggest that habitat fragmentation reduces global biodiversity but not necessarily local biodiversity (Schelhas and Greenberg, 1996; Laurance and Bierregaard, 1997). This seeming paradox occurs because some species benefit while others suffer from fragmentation. The result is that local communities change composition after fragmentation, but the overall number of species in a local area does not necessarily decline. Forest interior specialists and wide-ranging species, notably predators, may go extinct with extensive fragmentation, thus lowering global biodiversity. Generalist and exotic species typically increase in fragmented habitats. Research suggests that many forest species tolerate considerable fragmentation because they can utilize the surrounding matrix. The extent of their ability to sustain viable populations depends on resources being available in the matrix and similarity of the matrix to the remaining native habitat.

Research in Singapore and Hong Kong is beginning to reveal the long-term effects of fragmentation. Forest fragments on these islands have been isolated for 150 to 350 years. Extinction of birds and mammals has been severe, but many plants are able to persist. Forest interior plants that are sensitive to changes in temperature and humidity (such as orchids and epiphytes) are also extinguished by fragmentation.

C. Habitat Degradation

Changes in land use can affect biodiversity without changing land cover. Land cover can remain essentially the same while the intensity of use increases and degrades habitat. Three forms of land use that degrade land cover rather than drastically changing it are (1) agricultural intensification, (2) recreation, and (3) forest management for sustainable timber production.

1. Agricultural Intensification in Europe

Agriculture is the dominant form of land use in Europe. Conversion of forests to agricultural lands has resulted in dramatic changes in land cover. Forests once covered more than 80% of Europe, but today they cover only one-third of the land. In contrast, slightly more than 40% of the European landmass is now agricultural. Over one-third of this land is permanent grassland maintained for grazing; the remainder is arable land. Conversion of forest to agriculture is no longer a major conservation concern in portions of Europe like Britain that have few remaining primary forests. Instead, the intensification of agricultural practices is degrading the conservation value of lands with grave consequences for biodiversity.

Agricultural intensification in Europe comes in many guises (Newton, 1998). (1) Increased use of chemical pesticides and fertilizers. (2) Removal of hedges to increase field size. (3) Plowing in late summer after harvest rather than waiting until the following spring. (4) Draining wetlands to increase arable and pasture lands. (5) Conversion of mixed farms that produced a variety of animal and plant products to monoculture cereal farms. (6) Earlier harvest dates. (7) Intensive grassland management to increase grass growth. (8) Increased stocking density of sheep in hill country.

These changes in agricultural practices cause declines in bird distribution and density for three main reasons (Newton, 1998): (1) chemicals and early harvests reduce breeding success and survivorship of birds; (2) breeding and foraging habitat is reduced by hedge removal, land drainage, late summer plowing, intensive grassland management, and increased grazing; and (3) habitat diversity and resulting bird community diversity are reduced when mixed farms are converted to monocultures. The results of these changes have been dramatic. It is suspected that intensification has affected 42% of bird species of conservation concern in Europe. Nearly 90% of the 26 farmland birds in Britain decreased their range of occurrence from 1970 to 1990, and those species using pastures were negatively affected by increasing grazing intensity. Specific aspects of agricultural intensification affect particular species, but on the whole this form of habitat degradation threatens a significant portion of the common, as well as the rare, European avifauna.

2. Influence of Recreation on Wildlife Communities

As our landscape becomes more dominated by urban and agricultural land, our psychological need to visit and explore the last remaining wild places increases. We often fulfill this desire by camping, hiking, nature watching, hunting, and fishing in wildlands. These seemingly benign activities can affect biodiversity by harassing animals, trampling plants, and altering competitive, facilitative, and predator–prey relationships (Knight and Gutzwiller, 1995). Although recreation in wildlands rarely changes land cover, its effects are often greatest near settlements, thereby extending the actual area modified by settlement.

How campers and fishermen affect salmon-scavenging birds in the Pacific Northwest of North America is a good example of the subtle ways that our land use affects biodiversity without changing land cover. Three species of birds [American Crow, Common Raven (*Corvus corax*), and Bald Eagle (*Haliaeetus leucocephalus*)] vie for salmon that die after spawning each fall and winter. Bald Eagles are the largest and dominant scavenger, but they are also the least tolerant of human activity. Sport fishermen float and wade rivers in search of live salmon and often flush eagles from their meals. This disrupts eagle foraging and may lead to lower overwinter survival and reduced population viability. The rarity of eagles and their status as a national icon increase the desire of nature watchers to observe them during the winter, thereby disturbing them even further. Recreation effects do not stop here. Camping areas along rivers are common and they provide consistent sources of food that fuel and increase in the wildland crow abundance and distribution. Crows are tolerant of humans and quickly return to carcasses after disturbance. They attempt to eat carcasses dominated by eagles and ravens but rarely succeed until humans displace the eagles (and often ravens), thereby incidentally allowing crows to eat without the check of competitors.

3. Effects of Intensive Timber Management

Timber harvest directly affects biodiversity by (1) facilitating land cover change through clearing land for agriculture and settlement and (2) fragmenting contiguous primary reserves as discussed earlier. Intensive forest management to produce timber may also affect biodiversity in more subtle ways by (1) reducing the average age of the forest, (2) simplifying the age structure and diversity within forest stands, and (3) increasing the frequency of disturbance in forests.

An intensively managed forest is like an intensively managed agricultural field, with the focus on producing the most marketable timber in the least amount of time. However, the effects of timber management on biodiversity appear to be much less than the effects of agricultural intensification. Animals with large space requirements, especially requirements for the interiors of mature forests, do not fare well in managed forests. But animals and plants that utilize clearings, thickets, and young forests thrive in managed forests, which can actually increase local biodiversity by providing habitat for such species. Nonetheless, the loss of sensitive interior forest species usually leads to lower global biodiversity.

The long-term effects of intensive timber management on biodiversity are poorly known. We do not know if managed forests provide adequate resources for many birds during the nonbreeding season, nor do we know how the shifting nature of forest patches in time and space, so typical of managed forests, affects the ability of plants and animals to maintain viable populations connected by dispersal. It is safe to conclude that maintaining forested land cover is an important contribution to global conservation. The maintenance of managed forest, that perpetuate native forest trees and some of the biodiversity dependent on them certainly helps stem the global loss of native habitat to agriculture and settlement.

III. METHODOLOGICAL ISSUES IN THE ESTIMATION OF LAND COVER AND LAND USE

Much of the information presented in this article depends on the estimation of current land use and land cover on a global scale. Our conclusions depend directly on the quality of these estimates. Therefore it is important for the reader to have some insights into how these data are obtained and the degree of uncertainty they contain.

A. How Are Land Cover and Land Use Estimated?

Land cover is estimated directly by aerial photography at small scales and by satellite imagery at large scales. Aerial photographs resolve fine details in land cover at local or regional scales. Satellite imagery detects features over hundreds of kilometers, but small details are missed because the resolution is coarser.

An important satellite used for imaging the earth is Landsat. Its imaging abilities have improved over the years: current Landsat TM (thematic mapper) images have better resolution (30 m × 30 m) that permits the fine-scaled discrimination of land covers. Another satellite that is widely used is the AVHRR (Advanced Very High Resolution Radiometer). Its wider swath and coarser resolution (1.1 km) are good for describing large areas of vegetation.

While modern land cover can be estimated directly using photography and satellite imagery, land use cannot always be derived from the images. Hence statistics of land use supplement or substitute for imagery. The Food and Agriculture Organization of the United Nations (FAO) has reported tabulated data on land use and cover on an annual basis since the 1950s (these data are used in Fig. 6). The FAO does not gather data independently but ratheer compiles national data from returned questionnaires. Globally, the FAO is an important source for time-series estimates of land use and cover over the last half of this century.

B. Limitations in Estimating Land Cover and Land Use

Virtually every determination of land use and land cover is an estimate that includes some level of smoothing, averaging, and guessing. The general conclusions we make about how land use affects land cover and biodiversity are not affected by these limitations. However, the application of the maps and processes to specific areas on the planet may fail because of estimation.

An important limitation occurs in large-scale mapping of land use and cover. Some minimum mapping unit (the resolution) is used in all compilations of land cover and use. Only one use or cover can be assigned to this minimum unit, and it is the dominant use or cover that is assigned and assumed to occur uniformly over the entire unit. For example, a unit that is simply recorded as "forest" might actually be covered by 40% forest, 30% lake, 20% grassland, and 10% building. Recording only the dominant cover reduces the occurrence of dispersed (rarely dominant) covers that typify many human-dominated land covers. Therefore, large-scale maps likely underestimate the actual effect that humans have on land cover.

Projected early estimates of land use and cover are by necessity approximations. The map of preagricultural land cover (see Fig. 3) was derived entirely by modeling the plant growth suspected under varying climatological and soil conditions. Cropland estimates in the 1700s were estimated by assuming that 0.2 ha of land was cleared for each human (Richards, 1990). Even the population size at that time was an estimate. Regional estimates of land cover and use are also of inconsistent quality. FAO data vary greatly in quality by country because uniform scales and cover categories are not used. Yet even these large sources of estimation and inaccuracy do not affect the general conclusion that at a global scale agriculture has increased tremendously, replaced native forests and grasslands, and caused declines in biodiversity.

Another important bias is an underestimation of habitat degradation. Mapping and imagery capture the basic pattern of habitat conversion and fragmentation, but rarely detect changes in the quality of habitats that retain their basic structure. Selective logging, understory trampling and removal, and disturbance by human presence are not recorded on global scales. Regional, local, and more subjective measures are needed to assess these changes in land use.

See Also the Following Articles

AGRICULTURE, INDUSTRIALIZED • AGRICULTURE, SUSTAINABLE • DEFORESTATION AND LAND CLEARING • DESERTIFICATION • LANDSCAPE DIVERSITY • LOGGED FORESTS • RANGE ECOLOGY • TEMPERATURE GRASSLAND AND SHRUBLAND ECOSYSTEMS • TIMBER INDUSTRY

Bibliography

Berry, B. J. L. (1990). Urbanization. In *The Earth as Transformed by Human Action* (B. L. Turner, II, W. C. Clark, W. R. Kates, J. F. Richards, J. T. Matthews, and W. B. Meyer, eds.), pp. 103–119. Cambridge University Press, Cambridge, United Kingdom.

Collins, N. M., Sayer, J. A., and Whitmore, T. C. (eds.). (1991). *The Conservation Atlas of Tropical Forests: Asia and the Pacific.* Simon and Schuster, New York.

Houghton, R. A. (1994). The worldwide extent of land-use change. *Bioscience* 44, 305–313.

IUCN (International Union for the Conservation of Nature). (1996). *The 1996 IUCN Red List of Threatened Animals.* IUCN, Gland, Switzerland.

Knight, R. L., and Gutzwiller, K. J. (eds.). (1995). *Wildlife and Recreationists.* Island Press, Washington, D.C.

Kummer, D. M. (1992). *Deforestation in the Postwar Philippines.* University of Chicago Press, Chicago.

Kummer, D. M., and Turner, B. L., II. (1994). The human causes of deforestation in Southeast Asia. *Bioscience* 44, 323–328.

Laurance, W. F., and Bierregaard, R. O., Jr. (eds.). (1997). *Tropical Forest Remnants.* University of Chicago Press, Chicago.

Matson, P. A., Parton, W. J., Power, A. G., and Swift, M. J. (1997). Agricultural intensification and ecosystem properties. *Science* 277, 504–509.

Matthews, E. (1983). Global vegetation and land use: New high resolution data bases for climate studies. *J. Climate Appl. Meterol.* 22, 474–487.

Meyer, W. B., and Turner, B. L., II. (1992). Human population growth and global land-use/cover change. *Annu. Rev. Ecol. Systematics* 23, 39–61.

Meyer, W. B., and Turner, B. L., II. (eds.). (1994). *Changes in Land Use and Land Cover: A Global Perspective.* Cambridge University Press, Cambridge, United Kingdom.

Newton, I. (1998). Bird conservation problems resulting from agricultural intensification in Europe. In *Avian Conservation: Research and Management* (J. M. Marzluff and R. Sallabanks, eds.), pp. 307–322. Island Press, Washington, D.C.

Noble, I. R., and Dirzo, R. (1997). Forests as human-dominated ecosystems. *Science* 277, 522–525.

Richards, J. F. (1990). Land transformation. In *The Earth as Transformed by Human Action* (B. L. Turner, II, W. C. Clark, W. R.

Kates, J. F. Richards, J. T. Matthews, and W. B. Meyer, eds.), pp. 163–178. Cambridge University Press, Cambridge, United Kingdom.

Schelhas, J., and Greenberg, R. (eds.). (1996). *Forest Patches in Tropical Landscapes.* Island Press, Washington, D.C.

Skole, D., and Tucker, C. (1993). Tropical deforestation and habitat fragmentation in the Amazon: Satellite data from 1978 to 1988. *Science* 260, 1905–1910.

Turner, B. L., II, Clark, W. C., Kates, W. R., Richards, J. F., Matthews, J. T., and Meyer, W. B. (eds.). (1990). *The Earth as Transformed by Human Action.* Cambridge University Press, Cambridge, United Kingdom.

Vitousek, P. M., Ehrlich, P. R., Ehrlich, A. H., and Matson, P. A. (1986). Human appropriation of the products of photosynthesis. *Bioscience* 36, 368–373.

Vitousek, P. M., Mooney, H. A., Lubchenco, J., and Melillo, J. M. (1997). Human domination of earth's ecosystems. *Science* 277, 494–499.

LAND-USE PATTERNS, HISTORIC

Oliver Rackham
Corpus Christi College, University of Cambridge

GLOSSARY

climax The stable ecosystem that is supposed to result from a sufficiently long period of unchanging environment.

clone A group of individuals formed by vegetative reproduction from a parent. This term applies particularly to those trees and plants that reproduce by creeping stems or by suckers from roots, forming circular patches of genetically identical individuals.

coppice Regrowth of a felled tree from the stump.

cultural ecosystem Ecosystem produced by the long-term interaction of wild plants and animals with human activities.

endemic Plant or animal species limited to a small area of the world, for example, one island or mountain range.

heath Open vegetation dominated by dwarf undershrubs usually of the family Ericaceae.

meadow Grassland maintained by mowing.

moor Open vegetation, dominated especially by dwarf Ericaceae and *Sphagnum* mosses, on peaty soil.

pasture Grassland maintained by grazing domestic animals.

pollard A tree repeatedly cut at about 6–10 feet above ground to produce successive crops of wood.

savanna Grassland (or other nonforest vegetation) with scattered trees.

seminatural Vegetation that owes its character to human activity, but in which the plants are wild, not sown or planted. Examples: coppice woods; most heathland; some savannas.

wildwood Wholly natural forest not affected by sedentary human activities such as agriculture and pasturage, widespread in prehistoric times (including earlier interglacial periods) and still surviving in remote places.

woodland Forest forming part of a cultural landscape.

wood-pasture Cultural ecosystems, often savanna-like, combining trees and domestic animals.

SOME OF THE WORLD'S BIODIVERSITY is associated with "virgin forest" and other ecosystems that are supposed to have escaped human interference until now. This category shrinks as archaeological research reveals more of the extent and pervasiveness of past human activities. The landscapes of most of the world's land surface result from long and complex interactions between human activities and natural processes. Conservationists need to understand how such activities have shaped many landscapes that are now centers of biodiversity.

I. INTRODUCTION

Human intervention often does not efface all signs of previous ecosystems. The traveler in Texas, despite 160 years of often intense activity by settlers, has no difficulty in recognizing the presettlement vegetation regions (Pineywoods, Blackland Prairie, Post Oak Savanna, etc.).

Even in Britain, after 6500 years of settlement, the landscape is still not wholly artificial. By careful search, one still finds remains of the mid-Holocene vegetation categories (limewoods in lowland England, oakwoods in the north and west, pinewoods in east Scotland). Superimposed on these are the largely artificial ecosystems of arable farmland, sown grassland, and forestry plantations; the seminatural ecosystems produced by the responses of wild plants and animals to continuing land use (old pasture, hedges, moorland) or to past land use (abandoned mines, derelict forestry plantations); and the wealth of biodiversity contained in cities, old pits, and industrially derelict land.

Cultural ecosystems are often contrasted with wildwood or "virgin forest," the aboriginal vegetation of much of the world's land surface before the impact of settled human cultures. Wildwood itself, however, may already have been altered by the activities of nonsettled peoples. There is abundant evidence to show that hunting and gathering were not simple, unorganized activities, but could have profound effects on landscape and ecology, even in places remote from the activities themselves. An example is the extermination of elephants and other great mammals by Paleolithic peoples. The loss of the only animals capable of breaking down big trees could hardly fail to have profound, if as yet little known, consequences. From the early Holocene, if not earlier, land management by burning affected the ecology where vegetation was combustible, especially in Australia and North America. These activities, as far as we know, created only simple land-use patterns: they would alter, for example, the boundaries between forest and prairie or between combustible and noncombustible vegetation. Some of the Texas ecosystems, such as the Edwards Plateau savannas, were probably produced by Native American land use and taken over by settlers. The introduction of settlements, domestic animals, and cultivated plants—in European terms, the coming of Neolithic cultures within the past 8000 years—began to create the diversity and complexity of land-use patterns that is familiar today.

II. HUMAN ACTIVITIES AND CLIMATIC CHANGE

It used to be thought that, given enough time, the plant and animal communities in an area would settle down to a stable *climax* state, determined only by the climate and (maybe) geology and changing only slowly in response to evolution. This may have been so in the Tertiary geological period, but in the Quaternary the dominant factor has been climatic change. Glacial cycles in high latitudes destroyed all vegetation over large areas, including much of Britain. In middle latitudes and even in the tropics, although there has been little ice cover, climatic changes such as drought have been sufficient to displace or destroy most of the interglacial plant communities.

There have been many glacial cycles over the past 2 million years. In much of the world, forest has been the "normal" vegetation only in interglacial periods like the present, of at most a few tens of thousands of years each. Changes of climate have been too rapid for plants and animals, apart from annuals and other fast-reproducing species, to keep up with them by evolutionary change.

Human activity has recently become the predominant force shaping the world's ecosystems. However, it did not burst into a world of stable environment in which evolution had previously been the predominant force. It came on top of a period of unusual instability: climatic change had been operating on a time scale not much longer than that of human activity itself. In Britain humanized ecosystems have existed for about 6000 years; the "virgin forest" preceding them had existed for at most another 6000 years and was itself affected by the activities of Mesolithic and Paleolithic people.

Climatic change and human activity overlap in time, and their effects can be difficult to separate. In the Mediterranean, the present, rather arid, ecosystems are variously attributed to the strongly seasonal climate with its hot dry summers and to the activities of people "degrading" what would otherwise be a more forested landscape. In reality, the present climate there is only about 6000 years old and set in at the same time that Neolithic people were beginning to have a profound effect on the landscape.

FIGURE 1 Japanese cultural landscape of terraced rice fields and mountain forests. In the past hundred years, the higher terraces have been abandoned and replaced by forest. *Shugakuin, Kyoto*

III. PROPERTIES OF CROPS AND DOMESTIC ANIMALS

Land-use patterns depend on the ecological behavior of a few species of major crop plants and domestic animals, which are usually not indigenous to the region.

Most of the crop plants of European-style agriculture originated in the steppes of southwest and central Asia and have a particular set of ecological requirements. They have modest moisture needs, rather high needs for mineral nutrients, and are very intolerant of waterlogging and shade. As they spread, first to Europe and later to European colonies, farmers strove to convert more and more of the world's ecosystems to an imitation of a Turkish steppe in which wheat and barley are at home. Of the multitude of indigenous American crop plants, only a few—corn, potato, and tobacco—have been widely grown in Europe; these are the few whose ecological requirements are not too different from those of Southwest Asian crops, although they need more moisture. Rice is a Southeast Asian crop with very different requirements, growing in wetlands and making possible the great wetland civilizations of Asia. Hence it comes about that generations of farmers have sought to destroy Europe's fens and convert them into artificial steppe, whereas Japanese farmers, with the utmost ingenuity, have contrived to turn hillsides into artificial fen (Fig. 1).

Many tropical crops, on the other hand, will grow in moderate shade, and some require it. The distinction between forest and farmland, which is so sharp in most European-style agriculture, is not always clear in the tropics.

Keeping domestic animals also involves destroying or modifying natural ecosystems. Cattle, sheep, and goats eat tree leaves but are not forest animals. If kept in the shade of trees, they soon eat all the leafage within reach, creating a "browse-line" beneath which nothing is allowed to grow. To keep them in manageable numbers requires at least some gaps between the trees, in which grasses and low herbage will grow. Pigs, although often thought of as acorn feeders, cannot be kept in forest alone. Acorns are available for only a few months of the year and contain insufficient protein to sustain pigs by themselves. The world's cultural ecosystems would be very different had monkeys been domesticated for human food.

IV. FIELDS, HEDGES, AND TERRACES

In some parts of the world, arable fields are impermanent, being created out of forest or grassland, cultivated for a few years, and then allowed to revert to natural vegetation. The landscape consists of a mosaic of areas in different stages of reversion. This *shifting cultivation* is typical of regions too infertile to support even primitive agriculture permanently, especially in the tropics. In more fertile regions, such as most of Europe, people avoid having to repeat the immense labor of creating new farmland. Fields are permanent and tend to be private property with fixed boundaries.

In regions not yet reached by weed killers, arable fields are seminatural ecosystems. Many weeds came from the original homelands of the crops they accompanied, and some annuals have evolved in parallel with particular crops: for example, cow-wheat (*Melampyrum arvense*), whose seeds resemble a grain of wheat, so that they are sown along with the wheat crop. In Japan, rice fields are complex aquatic ecosystems involving cyanophytes, water ferns, amphibians, and insects; some of the plants fix nitrogen and so benefit the rice crop.

Fields tend to be discontinuous. As well as boundaries where they adjoin forest, pasture, or roads, they have boundaries separating ownerships or subdivisions within an ownership (Fig. 2). These are determined both by practicalities (e.g., erosion control) and by anthropological factors, such as land tenure, inheritance, or collectivization. The boundaries (hedges, banks, walls, etc.) provide a network of relatively natural ecosystems permeating the farmland. The size of fields varies from country to country. In parts of Canada or Australia, 100 acres is a small field. In Texas, the nineteenth-century allotments of 1 square mile (640

FIGURE 2 Anglo–Welsh cultural landscape of small irregular fields, thick hedges, many hedgerow trees, and scattered farmsteads, of unknown origin but likely to be well over 500 years old. *Llanveynoe, Herefordshire, England*

acres) are by now divided into at least 30 compartments in several ownerships. In the hedged parts of medieval England, a typical field would be of about 4 acres. In Crete, fields tend to be of about 1 acre. In Japan, with rice as an immensely high-yielding crop, even this would be a big field.

A *hedge* (in North America, a *fence-row*) is a row or strip of trees or shrubs forming the boundary of a field. In Europe, these are often maintained by cutting in particular styles to form a barrier to keep cattle or sheep from straying into the next field; often trees are left at intervals either uncut or pollarded (Fig. 3). In America, they are less managed.

In America and to some extent Europe, most hedges have arisen spontaneously as trees have taken root from wind- or bird-dispersed seed at the bases of fences or on balks or walls between fields. In England, hedges are thought of as deliberately planted, but can arise naturally, as many of the older hedges probably did. In Australia, hedges are rare, evidently because most eucalyptuses and other native trees have poor powers of seed dispersal.

FIGURE 3 Ancient hedge between two fields, managed by coppicing. It has been cut down (leaving a few trees standing) and is beginning to sprout. In 2 years it will be back to its normal state. *Polstead, Suffolk, England*

Hedges presumably go back in some countries to the beginnings of agriculture. In Europe, the earliest evidence for them is about 2000 years old; they probably go back much further, but leave little archaeological record. In England, they are well documented for about 1200 years. Many individual hedges are at least 500 years old; these tend to be mixed in their composition, unlike the monotonous hedges of more recent origin.

Hedges are often thought of as linear wood-lots. This is partly true: in England, many hedges were periodically felled as if they were woodland, yielding crops of fuel wood and timber. However, in England and (from what I have seen of them in America), hedges are distinct vegetation types in themselves, seldom corresponding in detail to the local forest. They usually lack the more exacting woodland plants, apart from those few hedges that are the surviving edges of grubbed-out wood-lots.

Hedges are important as diversifying what might otherwise be featureless farmland. They introduce tree-living (though not forest birds) to open country that would lack nesting sites. They are often regarded as corridors for movement of mammals across farmland.

In many mountainous and densely populated parts of the world, slopes are shaped into terraces to create flattish surfaces on which to grow crops (Fig. 4). Banks or walls between terraces, like hedges between fields, can be significant sites for seminatural vegetation.

V. FOREST AND WOODLAND

It is often supposed that wildwood consisted of trees, trees, and nothing but trees, and only shade-bearing ground vegetation; gaps arising through the death of trees would promptly be filled by the growth of new trees. Examples can be seen where the dominant trees are of shade-bearing kinds such as *Abies* and *Fagus* species. The argument that wildwood was like this is often based on the general doctrine of climax vegetation, rather than on evidence of the history of specific forests. Pollen analysis can exaggerate the dominance of trees, many of which produce more pollen than her-

FIGURE 4 Cretan cultural landscape in a semiarid climate. In the bottom of the basin are houses grouped into hamlets, with gardens, orchards, and small fields demarcated by dry-stone walls. Extending up the mountains are remains of terraces, on which grain used to be grown. *Asphéndou, Sphakia, Crete*

FIGURE 5 A coppice wood. The foreground has just been felled, leaving the stools from which new growth will arise. In the middle is an area of 2 years' growth since last felling. An older part of the wood forms the background. *Bradfield Woods, Suffolk, England*

baceous plants. If this is allowed for, some wildwoods were much more diverse.

In some tropical forests, there is vast biodiversity, owing to the many species of tree and the abundance of vines, epiphytes, and termites, which enable the trees themselves to support complex ecosystems. In other parts of the world, continuous forests, especially of densely shading trees, are relatively poor habitats: many of their plants and animals are concentrated in gaps, cliffs, watercourses, burnt areas, and other breaks in the continuous shade.

Ancient peoples destroyed some areas of forest to create farmland, grassland, and heath; they also affected the remaining forest by various forms of land management. Continuous forest is not very productive for most human purposes. Edible animals and plants occur sparsely, if at all, and the animals are difficult to catch. Tree fruits are out of reach in the high canopy.

Forests were not generally destroyed by people cutting down the trees in order to use them. I know of no instance in European history of a forest being destroyed—converted to nonforest—solely by people using up the trees. Normally they would cut the trees suitable for the purpose in hand and leave the other trees. The result would be a depleted forest, unsuitable for that purpose until a new generation of trees had grown. This is not to be confused with a destroyed forest.

Great trees do not easily furnish wood for fuel and timber for construction. A big tree, when cut down, is a very intractable object. Until the coming of sawmills, vehicles, and railroads capable of reducing giant trees, people preferred to use the smallest log that would serve the purpose and to manage forests to produce a succession of trees small enough to handle.

Coppicing is an important factor. Most European and North American trees, other than conifers, survive being cut down and sprout from the stump. Clonal trees, like European elms and American beech, sucker from the roots. Coppicing is one of the world's most important practices in historic forest management. By cutting down woodland every 5–30 years and allowing it to grow again from the stumps, a permanent succession of small stems can be assured, of sizes that are easily handled and suitable for light construction and fuel. It was often the practice to leave a scatter of trees of selected species to grow on for three or four cycles to yield constructional timber (Fig. 5).

Coppicing has been the nearly universal woodland management in Britain, well documented for the past thousand years and known on archaeological evidence for some 6500 years. It is the basis of historic forest management in many parts of the broad-leaved and Mediterranean zones of Europe and also in Japan. It was apparently not much practiced by Native Americans but was widely introduced by European settlers in America, where there are large areas of ex-coppiced wood-lots.

Coppicing affects biodiversity by drastically reducing the shade at the start of each cycle of felling and regrowth. Two or three years of relatively open conditions follow, ending as the new growth closes in. This favors various woodland plants and animals. Low-growing herbs such as species of *Viola* and *Primula* flower in abundance in the years of extra light (Fig. 6). Others

FIGURE 6 Oxlip (*Primula elatior*) flowering in abundance among hazel (*Corylus*) shoots of 1 year's growth since coppicing. *Hayley Wood, Cambridgeshire, England*

FIGURE 7 Eucalyptus forest 2 years after being logged for the third or fourth time. Note the abundant regrowth and ground vegetation. *Merimbula State Forest, New South Wales*

such as *Euphorbia* species appear from buried seed produced by their parents at the last felling. Many insects feed on the leaves or nectar of these plants. The middle stages of regrowth, when there is a thicket of young stems, favor warblers and similar small birds—a famous English example is the nightingale (*Luscinia megarhynchos*). The dormouse (*Muscardinus avellanarius*), an English woodland mammal, favors the later stages.

Coppicing is often thought to be artificial, but the ability to coppice is widespread among the world's trees and presumably was an adaptation to some process in wildwood. Sometimes it is a response to fire, but it is not correlated with flammability: few pines (among the world's most flammable trees) will coppice, but fireproof trees such as elms and poplars coppice or sucker. American, European, and Japanese species of *Tilia* (lime, basswood) are self-coppicing and grow naturally in a multistemmed form; so do American and Japanese species of *Magnolia*. Possibly coppicing behavior is an adaptation to tree-breaking mammals: the axes of woodcutters are a replacement of the missing elephants etc. I have observed something like a coppicing ecosystem developed in eucalyptus forests in New South Wales after only the third or fourth successive logging (Fig. 7).

Another historic woodland practice is the creation of permanent edges and open areas. In England, woodlots have permanent edges (many of them are over a thousand years old) defined by banks and ditches, constructed as a conservation measure (Fig. 8). They may also have permanent tracks and other open areas in the interior. In a typical wood lot, well over half the plant species are associated with the boundary, with recently felled areas, or with permanent openings. The species of permanent openings tend to constitute plant communities of their own, distinct from those of temporary clearings and of grassland away from woodland. On pollen evidence, something like these permanent open areas already existed in wildwood.

VI. WOOD-PASTURE AND SAVANNA

Where the environment becomes too dry or cold for forest or where there is too much grazing, forest changes into nonforest in various ways. The trees may suddenly stop, or there may be a zone of trees reduced to the stature of shrubs, or a mosaic of patches of forest and

FIGURE 8 Edge of an English wood. The bank with external ditch was constructed about 1000 years ago by the monks of Bury St. Edmunds Abbey to demarcate their wood-lot. The trees on the bank have just been cut: in the interior, they are of 1 year's growth. *Bradfield Woods, Suffolk, England*

FIGURE 9 Savanna in middle Texas. The prairie (rich in spring and summer flowers) is dotted with motts—clonal patches—of Texas live oak and Texas elm. This is a cultural landscape of Native American origin, which was already in place when the area was settled in 1846. It needs to be maintained by grazing or burning, without which it is invaded by juniper (*Juniperus ashei*) and turns into forest. *Valley Mills, Texas*

patches of nonforest. Alternatively, there may be a zone, often of great extent, of scattered trees among grassland. This constitutes tropical and subtropical savanna ecosystems and their extensions into northern latitudes. The trees may be single, as in the wood-pasture ecosystems of England, the *dehesa* landscapes of southwestern Spain, and the eucalyptus savannas of Australia. Or they may be in groups, as with the oak and elm motts of middle Texas (Fig. 9). This depends on the properties of the trees: Texan *Quercus fusiformis* and *Ulmus crassifolia* are clonal trees that sucker.

It used to be thought that the "natural" savannas of Africa and Australia were very distinct from the semiartificial treed grasslands and treed heaths of Europe, which were regarded as "degraded forest," meaning forest from which some trees had been removed. Recent research makes the distinction less clear: many tropical savannas are at least partly the result of ancient land management, especially by burning. The pollen record of European wildwood, moreover, often contains non-shade-bearing undershrubs and herbs, showing that it must have had open areas as well as trees.

Cultural treed grasslands are usually differentiated from forest by there being enough browsing animals to prevent trees from occupying all the space. The amount of browsing needed to do this varies. Where the climate is so dry or cold as to be marginal for tree growth, it takes less browsing than in regions, such as England, well within the climatic limit of forest. (In some hot-climate savannas, conversely, the grasses are the dominant partners, and too much browsing encourages trees.)

The limits of "natural" treed grassland are vague and controversial. The amount of browsing—the numbers and behavior of the animals—is the most difficult aspect of prehistoric ecosystems to ascertain. Trees tend to be exaggerated, because many nontree plants produce sparse or nondiagnostic pollen grains and also because pollen analysts are trained to interpret tree pollen in terms of forest rather than of single trees.

In glacial times, parts of Europe not covered by ice were evidently dry as well as cold. The natural vegetation consisted of either patches or scatters of trees, which pollen analysts call "forest steppe." This was the situation for at least four-fifths of the past 2 million years, punctuated by interglacials where large areas of forest became possible. In the present interglacial, when the climate became favorable for trees, non-shade-bearing plants became very local—except in the drier parts of southern Europe, where forest was still patchy and savanna-like.

When people took to keeping cattle, sheep, and goats, they replaced an unknown, but probably sparse, scatter of wild herbivores by local concentrations of domestic animals. To feed them, they turned forest into grassland; but they also kept domestic animals in the remaining forest. This involved managing the forest in a different way from where wood was the product. The trees would have to be sparse enough for pasture to grow between them on which the livestock might feed. Trees were also cropped, not by coppicing—for the animals would eat the young shoots and kill the stools—but by the practices of pollarding and shredding, cutting the tree high enough for the shoots to arise out of reach. Sometimes wood was the objective, but in climates where grass did not grow all year, it was the practice to harvest leafy tree branches to store instead of hay for feeding the animals in the cold of winter or drought of summer.

Such seminatural wood-pasture ecosystems came to occupy large areas of Europe and Asia, often associated with communal land management. In England, there was a further development with deer as semidomestic livestock from the eleventh century A.D. onward. Landowners would keep deer (native and introduced) for meat and as a status symbol in enclosed parks. The king, moreover, had the right to keep deer on certain commonlands, such as Epping Forest and the New Forest, and to kill and eat them. Parks and Forests often contained treed grassland on which deer and other livestock fed. There could also be elaborate systems of temporary enclosure to combine livestock with coppice woods.

FIGURE 10 A complex medieval wood-pasture, combining old grassland, ancient pollard trees, and coppice wood-lots: a particularly important juxtaposition of habitats, reproducing some characteristics of wildwood. *Hatfield Forest, Essex, England*

Wood-pasture is a most important ecosystem, especially for insects and birds that require both trees (for example as nest sites) and grassland (Fig. 10). The inland bird fauna of England as a whole falls into this category. Wood-pasture also can preserve ancient trees. A large proportion of the fauna and flora of trees, especially beetles and lichens, is associated with the specific habitats of old trees: old dry bark, heartwood rotted by particular fungi, and cavities of various sizes. These are specialized, poorly dispersed animals and plants and presumably existed in wildwood. Forest trees are not an ideal habitat, for the lives of old trees are often shortened by the competition of neighboring trees. Coppice woods and managed forests are inhospitable to them, unless the ancient bases of coppice stools are a suitable habitat. However, wood-pasture is often very good for old trees and the organisms that live on them. The trees are free from competition and can live to a great age, especially as pollarding prolongs the lives of many species. Most of the information so far comes from England, where the New Forest, Windsor Forest, and Sherwood Forest are major sites for the invertebrates of ancient trees.

Wood-pasture is now much reduced in northern latitudes, but I have found traces of it as far north as Sweden. The area surviving in England, though small, contains some supremely complex and important examples. It still flourishes in southern Europe, covering about one-sixth of Portugal and one-eighth of Spain, although few of these contain ancient trees. It extends eastward into the Himalayas. At one time, it was evidently extensive in Japan, where old pollard trees can still be found embedded in forests. In North America, Native Americans and settlers both operated savanna-like landscapes, but pollarding was rare. In Australia, savannas are widespread and sometimes contain trees dating from presettlement times, but often the grassland component has been displaced by introduced grasses and herbs.

Wood-pasture is an important ecosystem for biodiversity: it perpetuates a different set of plants and animals from coppice woods. The most significant examples remaining are those few in which ancient trees and seminatural grassland are both preserved.

VII. GRASSLAND

Many of the world's grasslands are regarded as natural prairies, in regions too dry, or too cold, or too much grazed for trees. Others are part of a cultural landscape in regions that would naturally be forested; if not grazed, they revert to forest. There are large intermediate areas where natural grassland has been extended by burning and other forms of land management.

Much grassland in Europe, America, and Australia is sown pasture, little different from other arable crops. This practice goes back only three centuries. Previously, grassland was permanent, managed as either pasture (grazed) or meadow (mown for hay). Pasturage often goes with communal land use, as in the American Wild West or the chalk downlands of England.

Seminatural pasture grassland in England goes back to the Neolithic period, when the forests that covered the chalklands were converted into open land; prehistoric monuments such as Stonehenge were meant to be visible from a distance. Often there was a phase of farmland before the historic grasslands were established.

Meadow is a more productive kind of grassland, yielding hay which is dried and stored to feed animals in winter when the grass is not growing. This requires iron tools with which to cut the grass. As a large-scale land use, meadow is mainly of the past 1500 years. It was located especially in floodplains, which in medieval England were the most valuable land. In the sixteenth century, elaborate irrigation systems were developed to advance the growth of meadow in spring. Grasslands were often part of a highly integrated pattern of land uses that transferred nutrients, via the dung of cattle and sheep, to plowland and meadow.

A contrasting type of hay meadow is that on mountains, for example, in northern England or in the Alps. In the northern Apennines, the grasslands, originating in the Iron Age, were maintained for centuries to feed

FIGURE 11 Heathland, several hundred years old, now insufficiently grazed: if trees are allowed to take it over, they will destroy its value as a habitat. *Holt, Norfolk, England*

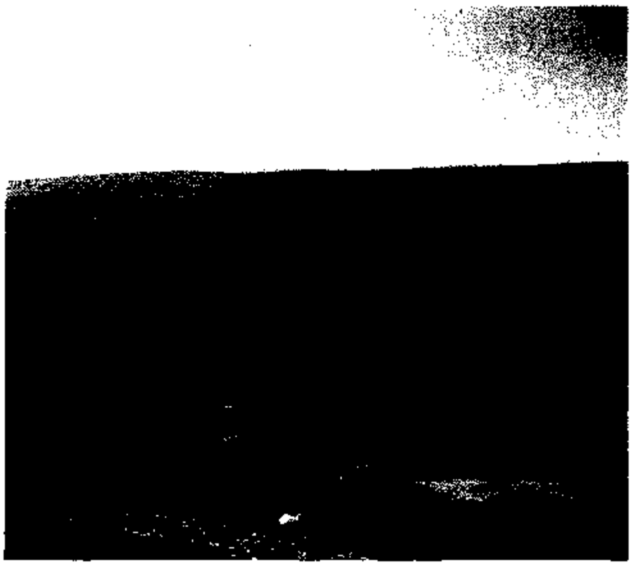


FIGURE 12 Medieval landscape on the border of Wales, with hedged fields and farmsteads in the valley and moorland on the thin soils of the plateau and on screes below the cliffs. *Cwm Olchon, Herefordshire, England*

local livestock and those of the city of Genoa; recently, they have largely turned into forest and their distinctive plant communities have been lost.

Old grasslands are the classic "seminatural" vegetation. They are artificial in that they were created by human activity, and disappear if that activity is withdrawn. They are natural in that nobody sowed the grasses and other plants of which they are composed. They are of many kinds, depending on whether they were pasture or meadow, on the geology, and on which animals fed on them. Chalk pasture is a dry, relatively infertile form of grassland that is one of the richest plant communities (in terms of plant species per square meter) in Britain. When grazed by sheep or rabbits, it is particularly rich in dwarf herbs and small undershrubs. At the other extreme, wet meadow is rich in tall grasses, sedges, and orchids.

VIII. HEATHLAND

Heathland in northwestern Europe is open vegetation other than grassland, dominated by dwarf undershrubs especially of the family Ericaceae. It can be a very rich plant community, with many herbs and also characteristic butterflies and birds. It is a seminatural ecosystem, maintained by grazing and cutting; without these, it quickly turns into woodland (Fig. 11).

Heathland began in a small way in the Mesolithic period, about 10,000 years ago; people may have encouraged it as part of land management in favor of deer. It expanded in later prehistory, especially as browsing by domestic livestock encroached on forest. Throughout the historic period, it was extensive in England, especially on acid soils in low-rainfall areas.

Heathland should perhaps be seen as a northward extension of the undershrubby plant communities of the Mediterranean, known as *garrigue* and *phrygana*. These, too, are often thought of as the result of destruction of forest: on palynological evidence, they were relatively local and specialized ecosystems until the Neolithic period. However, they are more stable than heath, occurring in places too dry for tree growth. Often they fail to turn into forest if grazing is removed. Phrygana is one of the richest ecosystems in endemic plants.

The heaths of northwestern Europe have fared very badly in the past 200 years. Little remains outside Britain, and even there, heaths, where not made into arable fields or forestry plantations, have been allowed to turn into woodland.

IX. MOORLAND

Moorland, although often confused with heath, is really a southward extension of the peaty tundra of the Arctic. It is characteristic of northwestern Europe, especially mountains (Fig. 12); it still covers half of Scotland.

Moorland is dominated usually by ericaceous undershrubs, especially heather, *Calluna vulgaris*. Unlike heath, it occurs in high-rainfall areas and has a peaty substratum. It grades into blanket bog, with pools and abundant *Sphagnum*.

Moorland is more stable than heath and less obviously a cultural ecosystem. It gradually extended during the Holocene, covering areas that once had forest. Sometimes there was an intermediate stage of pasture or even arable land: on the plateau of Dartmoor, southwest England, the boundaries of Bronze Age fields can still be traced. However, it is not clear how far the spread of moorland was due to the lapse of time and accumulation of peat, to a change to a wetter climate encouraging peat to grow, or to human activities tipping the balance against forest.

In historic times, moorland has been much used for pasture and for digging peat as fuel. Large supplies of peat have made up for the scarcity of wood fuel in Scotland and Ireland. A skilled tradition of periodic burning has grown up to maintain the pasture. In the past 200 years, moorland regions have lost human populations, but without forest returning. The main uses of moorland now are for grouse (*Lagopus scoticus*) and deer as game animals.

X. BUILDINGS AND BUILT-UP AREAS

Many plants and animals live on buildings, including some quite rare species: for example, the stork in Europe, which nests almost exclusively on buildings and structures, sometimes on platforms provided for it. In England, buildings and abandoned mines are the chief strongholds of bats.

Cities are often the chief concentrations of wildlife in their regions: well-known examples are Cambridge and Glasgow, which contain more wild plant species in a given area than anywhere else in their surroundings. This is because of the diverse habitats provided by gardens (despite the prevalent use of garden chemicals), buildings, and especially ruins and derelict land.

Derelict industrial land can be an important habitat. In England, the quarry that provided the excellent Barnack building stone is now one of the best examples of limestone grassland. Long-standing chemical contamination can produce its own ecosystems, especially where heavy-metal-tolerant plants have evolved.

Sacred sites develop in peculiar ways. In Japan, temples and shrines often own forests, which may preserve ancient trees and evidence of past management practices. In England, many churchyards are excellent habitats (Fig. 13). They often contain their own peculiar types of grassland. Tombstones and the stonework of the church itself are important sites for rare lichens, especially in areas that lack natural rock outcrops.

FIGURE 13 Churchyards have a range of habitats from old grassland (important for reptiles and amphibians) to the stone and brick surfaces of the boundary wall, tombstones, and the church itself (lichens) and the thatched or tiled roof (bryophytes, flowering plants). *Rockland St. Peter, Norfolk*

XI. THE RISE AND FALL OF HISTORIC LAND-USE PATTERNS

Human intervention creates new ecosystems as well as destroying old ones. Cultural ecosystems are often complex, diverse, and stable. Sometimes they are modifications and extensions of ancient natural ecosystems. Chalk grassland, for example, is in some respects a prolongation in time of the grasslands of the glacial periods. Moorland is in some respects an extension of tundra into southern latitudes. Some, however, seem to be new creations: I hesitate to guess what the habitat of the plant *Fritillaria meleagris* might have been before hay meadows had been invented.

Some conservationists regard ancient human intervention as destructive, like the modern logging of the Oregon redwoods or the early-modern ravaging of oceanic islands. Sometimes this was so: we would undoubtedly now disapprove of the destruction of almost all the native mammals of Crete by its early human inhabitants. Some interventions were experimental and unsustainable, like the prehistoric attempt to farm Dartmoor. Others may have been almost too gradual to notice, such as the erosion of tree cover on a common pasture.

The ecosystems of cultural landscapes could be preserved for centuries, both by ownership patterns and by deliberate conservation. By A.D. 1350, some 94% of the forest of England had been converted to nonforest uses. Woodland management had been invented long before, probably for other reasons, but as forest became scarce, conservation practices gradually spread to the remaining forest. It may be no coincidence that the idea of sustained yield is first set down in writing in 1356. Many medieval wood-lots still exist, and are often now nature reserves: "Ancient Woodland" is a scheduled category in modern conservation.

Historic land-use patterns usually worked on a small enough scale to result in useful juxtapositions of habitats. They often perpetuated habitats for non-shade-tolerant plants and animals. Because of the glacial history of the past 2 million years, many species are adapted to nonforest conditions and are threatened by the advance of forests in interglacial periods. They would have fared particularly badly in the present interglacial, because of the widespread extinction of giant tree-breaking mammals. For example, the greatest concentration of endemic plants in Europe is in the Mediterranean, where nearly all of them are nonforest species.

Historic land-use patterns have been in retreat for some 250 years. Usually this was because of outside ideas and interventions, not because they were unsustainable. Commonlands were not destroyed because of the operation of the "Tragedy of the Commons:" the supposed tendency of communal systems to break down because each participant pursues his own short-term advantage regardless of the rights of other participants. That had already been foreseen, and the participants in most commons had drawn up rules of use to prevent it.

In the eighteenth century, multiple and especially communal land uses became unfashionable. The only proper use of cultivable land (or land thought to be cultivable) was conventional agriculture, carried on by private owners in rectangular plots of not less than 10 acres. For uncultivable land the proper use was timber production organized by the state on the model of either French or German "scientific" forestry (depending on which state). Agricultural writers invented these ideas and bent the ear of governments to get them put into practice; they were not deterred if their schemes failed to work. In this way, the open fields of England were destroyed, and most of the heathland was converted into poor-quality arable. In countries where modern forestry was influential, such as Sweden, Germany, and Italy, most of the wood-pasture ecosystems were destroyed.

In the twentieth century there have been four contrasting processes: extension of cultivation (including plantation forestry) into areas not previously cultivated; intensification of cultivation in areas already cultivated; urban development; and land abandonment. Often, as in Crete, all four occur almost side by side. They destroy existing seminatural ecosystems without creating an equivalent. A common feature is that the grain of the landscape is coarsened: the small-scale pattern of juxtaposed land uses is replaced by monotonous expanses of the same land use.

One consequence is loss of habitat for creatures that require more than one habitat: birds that feed in the open but nest in tree holes; insects whose larvae feed in rotten wood but whose adults need a nectar source; plants that are weakly competitive when growing in the open but do not flower in shade—all these were favored by the mosaic of historical land uses, but not by what has replaced them.

Land abandonment does not re-create wildwood (wildwood either as it was before human intervention or as it would be by now had that intervention never happened). It tends to produce uniform expanses of even-aged, densely shading trees: the herbaceous plants of grassland, field edges, etc. disappear without being replaced by woodland herbs. In Mediterranean Europe, where secondary forest often consists of fire-adapted trees, the effect is a recurrent cycle of fires, which destroy whatever escaped the increasing shade. Many conservationists disapprove of goats, but the consequence of removing goats is usually to convert a browsing-dominated landscape into a fire-dominated one (Fig. 14).

Land use has become polarized: for example, land is now either forest or pasture but not both. Most of rural Japan, for example, is forest, and the remainder is intensively cultivated rice fields. The forests are now little used, and within them are the remains of many

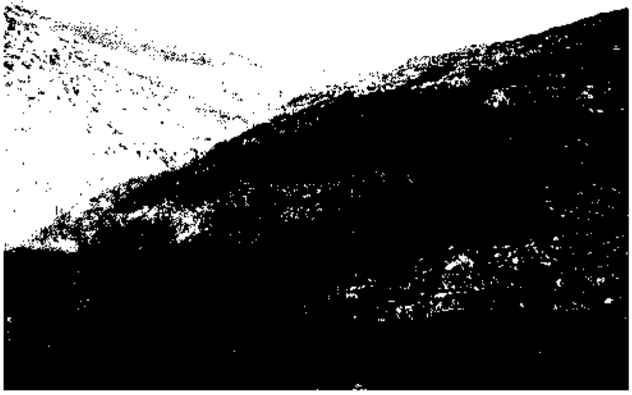


FIGURE 14 This was a goat-browsed landscape with patches of evergreen oak savanna and forest. It was taken over by foresters, who suppressed goats and planted pines (an ultraflammable tree). The inevitable result was a great fire. *Valor, Alpujarra, Spain*

historic ecosystems: coppice woods, wood-pastures (Fig. 15), terraced rice fields, pine savannas, Japanese early-modern forestry, and the growing of the giant grass with which house roofs were thatched.

Forests, on the whole, have become denser as savannas infill, coppice woods grow up, and foresters encourage the growth of timber trees. The older trees in a forest are commonly more spreading than the younger ones, having grown up when they had more room. I have seen examples in many countries from Canada to New South Wales, often in forests with little obvious human intervention.

An important innovation is the cult of tidiness: the urge among professional planners to destroy old quarries, spoil heaps, ruined buildings, patches of roughland, and even old trees and neglected tombstones, on the grounds that the public might not like them—even though, in practice, derelict land is often a much-loved public amenity.

FIGURE 15 An old pollard larch (*Larix leptolepis*) in what is now forest. Its spreading habit proclaims that it was originally a freestanding tree in grassland. *Tateshima, Chino, Japan*

XII. IMPLICATIONS FOR BIOLOGICAL CONSERVATION

Some conservationists disdain cultural ecosystems on the grounds that they result from the "degradation" of wilderness ecosystems, especially forest. The only proper business of biological conservation (they say) is to preserve pristine natural ecosystems where they still exist or to restore them where they do not.

However, this cannot be the whole purpose of conservation. In many parts of the world, such as England, ecosystems have been modified by human activity for much of the Holocene; not enough is known about pristine natural ecosystems to make restoring them a practical objective. Conservation depends on public interest, which it cannot retain if it is limited to remote wildernesses that few of the public ever see. It is unrealistic and impolitic to cut off biological conservation from other kinds of conservation: to separate the archaeological interest in an earthwork from the biological interest in the vegetation growing on it. It is essential for students to gain a knowledge of how local ecosystems work before pronouncing on those of distant countries.

Wilderness philosophy often ignores the realities of archaeology or vegetation history. In many countries, when a national park is scheduled, the authorities automatically expel the human inhabitants (however long and respectable their history), play down their part in the development of the landscape and ecosystems, and pretend that the park is wilderness. In Yellowstone National Park they even tried to efface all evidence that settlement had ever happened. In reality, few national parks lack human influence altogether; many of them are the last strongholds of precisely those old-fashioned land uses that biological conservationists ought to support. (The National Trust, biggest landowner in the Lake District National Park in England, goes to much trouble to sustain farming and prevent land abandonment.)

Conservationists are tempted to start from theories of what natural ecosystems ought to be and to re-create them in the image of those theories. They regard large, continuous, mixed-aged forests of tall trees as natural and try to preserve those forests that most nearly approach that ideal. The notion owes more to climax theory and to foresters' notions of an ideal forest than to weighing the evidence of what particular natural forests were like; yet its influence leads to neglect of discontinuous forests, savanna, and forests of short trees, however remarkable.

Biological conservation has many aspects. Priority is rightly given to preventing logging of remote forests that have never been logged, whether or not later research shows that they are really wilderness. However, the real merits of cultural ecosystems should not be overlooked. Wildwood included a great diversity of ecosystems. Those that involved open ground often predominated during glaciations. When giant herbivores were exterminated, this probably made the forests more shady, and the survival of non-shade-bearing plants through interglacials more difficult. Historic land management, to some extent, supplied the place of the missing elephants and restored the habitats of savanna and open ground. Modern changes tend to destroy historic ecosystems altogether or to replace them by uniform, very shady forest. Historic land-use patterns need to be respected and historic cultural ecosystems maintained. Conservationists rightly campaign against destroying trees in tropical rain forests. They often need to campaign against allowing trees to grow in unsuitable places near home.

Acknowledgment

Some of the plates are the result of travels with Mr. A. T. Grove, Dr. D. Lunney, Professor J. A. Moody, Professor Jun-ichi Ogura, Dr. J. Sleath, Professor Kathuyoshi Tsuchida, and the Sphakia Survey.

See Also the Following Articles

AGRICULTURE, TRADITIONAL • CATTLE, SHEEP, AND GOATS, ECOLOGICAL ROLE OF • ECOLOGY OF AGRICULTURE • FOREST ECOLOGY • HUNTER-GATHERER SOCIETIES, ECOLOGICAL IMPACT OF • INDIGENOUS PEOPLES, BIODIVERSITY AND • TRADITIONAL CONSERVATION PRACTICES

Bibliography

Birks, H. H., Birks, H. J. B., Kaland, P. E., and Moe, D. (Eds.) (1988). *The Cultural Landscape: Past, Present and Future*. Cambridge Univ. Press, Cambridge, UK.

Buckley, G. P. (1992). *Ecology and Management of Coppice Woodlands*. Chapman & Hall, London.

Cronon, W. (1983). *Changes in the Land: Indians, Colonists, and the Ecology of New England*. Hill & Wang, New York.

Dargavel, J. (1995). *Fashioning Australia's Forests*. Oxford Univ. Press, London.

Grove, A. T., and Rackham, O. (2000). *The Nature of Mediterranean Europe*. Yale Univ. Press, New Haven, CT.

Moreno, D. (1990). *Dal Documento al Terreno*. Il Mulino, Bologna.

Naveh, Z., and Liebermann, A. (1994). *Landscape Ecology: Theory and Application*, 2nd ed. Springer, New York.

Parra, F. (1990). *La Dehesa y el Olivar*. Ediciónes del Prado, Madrid.

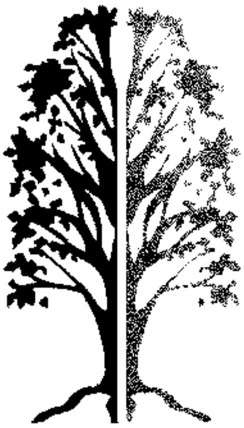
Rackham, O. (1986). *The History of the [British and Irish] Countryside*. Dent, London.

Rackham, O. (1990). *Trees and Woodland in the British Landscape*, 2nd ed. Dent, London.

Rackham, O. (1994). *The Illustrated History of the Countryside*. Weidenfeld & Nicholson, London.

Rackham, O., and Moody, J. (1996). *The Making of the Cretan Landscape*. Manchester Univ. Press, Manchester.

Vos, W., and Stortelder, A. (1992). *Vanishing Tuscan Landscapes*. Pudoc, Wageningen.



LATENT EXTINCTION—THE LIVING DEAD

Daniel H. Janzen
University of Pennsylvania

GLOSSARY

agroscape The agricultural, ranching, and plantation countryside, with its roads, irrigation ditches, buildings, and so on. The agroscape stands in contrast to the wildland countryside that is not directly managed by humanity (though it is strongly impacted by it). The agroscape intergrades with wildlands in the form of woodlots, abandoned fields, poor soil sites, hedgerows, and edges of wildlands.

living dead An individual stripped of the ecological circumstances that allow it to be a reproductive member of its population, but which is living out its physiological life. Living dead are most easily observed as large trees remaining on the agroscape, but they are also present in natural ecosystems.

megafauna Large mammals that are wolf-sized, deer-sized, and larger. Commonly used in reference to the many species of extinct "Pleistocene megafauna" that 9000 years ago populated the New World. The elimination of this megafauna by hunting (of the herbivores) and starvation (of the herbivore-deprived carnivores) was probably the first, and certainly the most dramatically irreversible, of the anthropogenic macroalterations of New World ecosystems. Today, of the extinct Pleistocene megafauna, only the horse remains—evolutionarily invented in the New World but surviving in the Old World until brought back as a gift from the Pleistocene by Spanish soldiers.

TREES AND MANY OTHER organisms that dot the tropical agroscape are often living dead. These are those individuals that have been stripped of the ecological circumstances that allowed them to be a reproductive member of their populations but are living out a physiological life. The term may also be applied to a portion of a population or a patch of vegetation. There are degrees of "living deadness." A living dead individual or even population may be resuscitated through ecosystem restoration. Ecosystem alteration by humans frequently produces living dead, but living dead are also part of natural ecosystem structure. The term and concept are conveniently applied to individuals that live long enough or are conspicuous enough to be included in the lay perception of the environment. The living dead are, in their sum, a latent extinction of a species in a place. This renders them a perceptual problem in the

Encyclopedia of Biodiversity, Volume 3

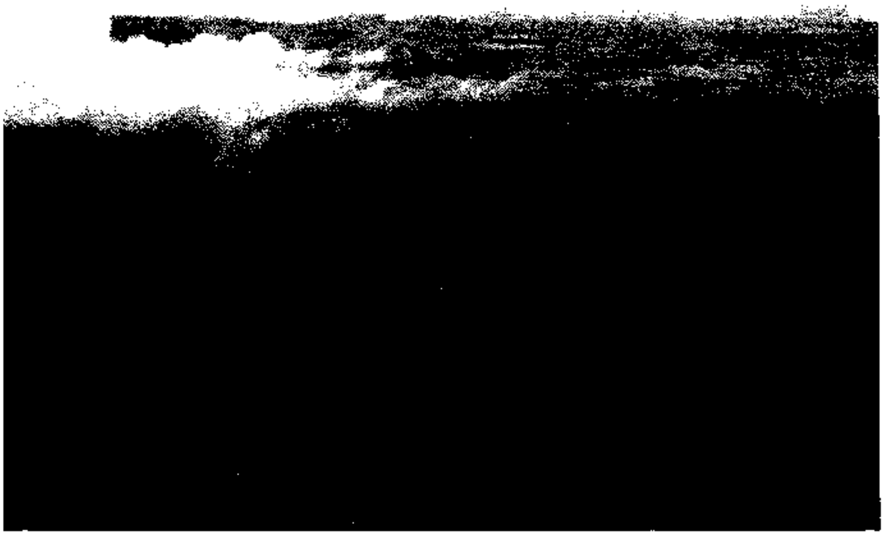


FIGURE 1 Living dead trees isolated in pasture at the edge of the agroscape (background) as it creeps into old growth forest (foreground). Los Naranjos, Sector Cacao, Area de Conservación Guanacaste, July 29, 1987.

psychology of tropical conservation because their presence obfuscates pending extinction. But living dead are also primary elements of natural processes of local extinction, immigration, and population-community structural dynamics in response to short- and long-term environmental change, be it natural or anthropogenic. Extra-tropical habitats and extreme tropical ecosystems may have fewer living dead than do complex tropical ecosystems, but they are nevertheless present. As magnificent as the living dead may be on the tropical countryside, I suggest that we not be distracted by attempting to save them, but rather that we focus our conservation efforts on saving large blocks of wildland ecosystems that are relatively complete and (it is hoped) relatively poor in living dead.

I. INTRODUCTION

The idea of the living dead has gradually emerged in my ecological understanding as I have lived past and around the majestic forest giants left standing as the agroscape creeps into Costa Rica's forest ecosystems over the past 4 decades (Fig. 1 and Janzen, 1986a, 1986b). This creep gradually converts the forest to an agroscape of pastures, fields, and roadsides dotted with the occasional adult tree but few or no juveniles. This is an agroscape where a magnificent flower crop now stands bee-less, an agroscape where fruit crops lie rotting below the pasture tree, an agroscape where tree seedlings wither in the dry-season sun or are turned to smoke in the dry-season anthropogenic fires.

I begin this article with a focus on adult large trees and use familiar examples from the Costa Rican countryside. To create breadth, I suggest that you join these verbs with the nouns from the ecosystems you know. This is a conservation biology question, but it applies to more than that, and it applies across the once-forested tropics as well as elsewhere.

Looking across the tropical landscape, the eye is greeted by stately single trees (Fig. 2), by patches of forest, by the blaze of a colorful flowering episode. Put an inventory to the plant species in a field, in a valley, in an ecosystem. All these species appear in the list. All is more or less well, we conclude, as 96.4% of the species that were here 50 years ago are still present. But are they? How many of them are living dead, part and parcel of latent extinctions?

We live a perceptual lie as we bustle about our agroscapes. That single stately green *Dipteryx* or *Hymenaea* or *Swietenia* or *Enterolobium*, standing in a field, pasture, or roadside, is often just as dead as if it were a log in the litter or the back of a logging truck. That tree was birthed in some favorable circumstance, a circumstance for pollination, seed dispersal, seed germination, and sapling survival.

FIGURE 2 A living dead *Terminalia* tree stands in silhouette, left behind as the rain forest was cleared around it, the natural tree falls in which its seedlings might have survived long since removed. Rincon Rainforest, Area de Conservación Guanacaste, January 6, 2000.

But one or more of these circumstances is now gone. It was carried away with the forest, put on the hunter's table, pesticided out of existence, or global warmed into oblivion. The long-lived tough adult lives out its physiological life, in the absence of the carpenter with a chain saw, but it is evolutionarily dead. Its pollen no longer flows to other members of the population, its seeds are no longer carried away from seed predators, or its seeds are no longer carried to a favorable site for seedling growth and sapling survival to adulthood.

But because the adult lives on, we are lulled into thinking that the environmental damage really is not all that bad, that extinction has not already occurred. If we can still show the tree to our children, it seems not to be extinct. It is so big and green and strong. Every year we see its flowers, and maybe we even see its fruits on the ground below. And after all, it has clearly weathered all that we have thrown at it. What ever can the matter be?

Humanity's interaction with the world's ecosystems has an enormous perceptual element. We act on what we perceive, be it threat or opportunity. Much of our conservation pragmatics and understanding is based on our knowledge that we really are losing species, losing ecosystems, losing the capacity of the environment to absorb our footprints. But that knowledge comes from what we see and measure. If all members of a tree species were to have the trait that each abruptly falls over dead the moment that it ceases to be a reproductive member of its population in its ecosystem, there would be far stronger alarm cries across the tropics about extinction rates and realities. If trees, the largest organisms on most of our landscapes, were very short lived as compared with humans, there would be less of perceptual problem—though just as large a conservation problem.

When the terrestrial world was covered with forest ecosystems, the single tree left standing in an aboriginal cornfield may well have been living dead, but the population from which it was derived was not usually at risk of anthropogenic extinction, unless perhaps dependent on a seed disperser targeted by that aboriginal population (Janzen and Martin, 1982). But when the agroscape is dotted with living dead in the wake of contemporary omnipresent ecosystem alteration, latent extinction is very real. A tree species may be ranked as "common"—meaning visible from a car window along many roads—yet be effectively extinct in a county, state, or region. And since the agroscape now stretches from horizon to horizon, the plant may well be absolutely extinct, since all of its former range may be populated by living dead.

II. DEFORESTATION AND THE LIVING DEAD

The forest need not be removed to convert trees to living dead. It is just that when the forest is partly removed, there is a very high chance that this alone will ecologically deprive many individuals of the remaining tree species sufficiently to convert them to living dead status. And, it certainly leaves the living dead very visible.

But even when the forest is left in place, that is no guarantee of a healthy tree population. When the Pleistocene hunters and their carnivorous helpers hunted out the neotropical mastodons and gomphotheres, the glyptodonts and camels, the ground sloths (Janzen, 1983b; Janzen and Martin, 1982), they did not do it by forest clearing. For decades to millennia after this 9000-year-old event, many of the remnant individuals of the tree populations that these big mammals

FIGURE 3 A living dead *Crescentia alata* fruit crop presented to earthbound extinct megafauna (Fig. 4). Sector Poco Sol, Area de Conservación Guanacaste, May 28, 1988.

dispersed (Fig. 3), and for which they created safe sites for seedlings by their browsing and trampling, would have been living dead scattered in the forest.

If some particular species—a pollinator or dispersal agent, for example—in the forest is extinguished, by whatever cause, there will often be surrogates and alternates that will assume, in some form, some portion of the "role" of the extinguished mutualist. The tree species will live on, albeit in some other ecological morph, and therefore in some technical sense will not be extinct. The tree that was "dependent" on the extinguished species will not, then, be living dead. But the devil is in the details. We need to go case by case. The suite of interactants with a tree species generates a given seed shadow, pollen rain, sapling demography, and microgeographic distribution. Remove one species of interactant. The entire n-dimensional hyperspace shifts in this or that direction. In some places this is toward eventual extinction, in other places it is just a change in demography and microgeographic distribution.

The history of any surviving species is that it must have survived thousands of such handoffs from one mutualist to another, from one moment to the next (e.g., Hallwachs, 1986). What bumps individuals into the category of living dead is the serendipitous event of losing irreplaceable partners. Humanity has a way of removing not only partners, but whole suites of them, as well as altering the physical environment. Our thoroughness and omnipresence creates ecological irreplaceability. Yes, when we lose one ground sloth, a glyptodont picks up some of the slack, though the tree is now a different beast. And at some time, likely as not, some new slothoid arrives by evolution or immigration over the millennia. But lose all these big mammals at once, and the result is guaranteed to be large arrays of living dead.

We have all been nourished by the marvels of evolutionary understanding, leading to the temptation to wonder if rapid evolution will not resuscitate a living dead population, if not many of its individuals. Novel pollinators, dispersal agents, fruit morphology, flowering phenology—all could save the day. In theory yes, but in reality not on the timescales ordained by humanity's charge across the landscape. How long will it take to evolutionarily reinvent a neotropical herbivorous/frugivorous megafauna? Fracture the remaining forest, with its living dead, into small ecological islands (also known as national parks and reserves). Thereby create ideal circumstances for rapid and novel evolution. We still cannot expect natural selection to create a mastodon from a white-tailed deer in anything like the speed required to be an antidote for neotropical rain forest anthropogenic alteration, beginning with the megafauna extinctions.

Certain kinds of habitat destruction are compatible with some tree natural histories. Two common trees, the guanacaste (*Enterolobium cyclocarpum*, Fabaceae) and jicaro (*Crescentia alata*, Bignoniaceae), owe their

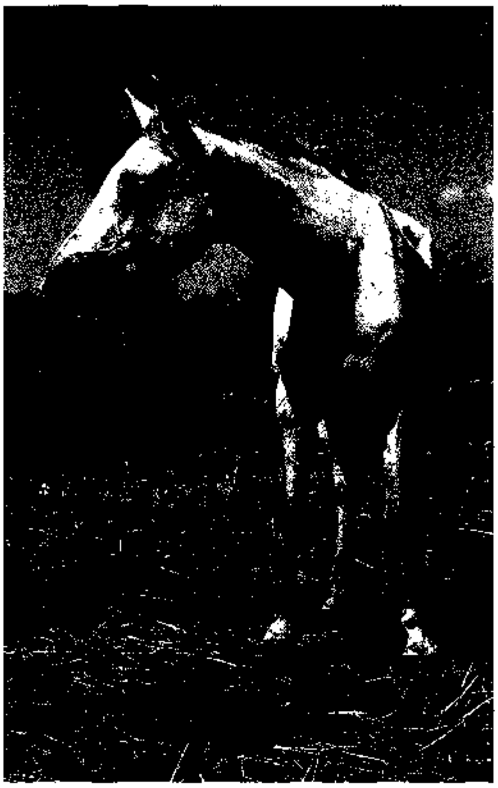


FIGURE 4 An earthbound extinct megafauna returned from the Costa Rican Pleistocene by Spanish immigrants, breaking a *Crescentia alata* fruit (Fig. 3) to eat the molasses and seeds inside. Sector Santa Rosa, Area de Conservación Guanacaste, 1980.

contemporary prominance on the Mesoamerican Pacific coastal landscape to a particular kind of habitat destruction. For both, large mammals—such as free-ranging horses—swallow the seeds while eating the content of indehiscent fruits fallen below the parent tree (Figs. 3 and 4), and later defecate them in open sunny habitats (Janzen, 1981, 1982a, 1982b). Forest clearing unto brushy pastures and scraggly roadsides, populated by widely circulating working horses, maintains a healthy population of reproducing guanacaste and jicaro trees in a precarious balance with humanity.

What did these trees do before the Spaniards brought the horse back from its Old World refuge after its neotropical extinction by Pleistocene hunters (Janzen and Martin, 1983)? They probably survived in a peculiar habitat characterized by ample insolated ground yet sufficient rain for there to be large trees and sloppy seed predator rodents (or human fruit and seed harvesters), which offered sufficient seed dispersal. River edges, marsh edges, and the interface between tropical dry forest and desert are such habitats, and the aboriginal village/field edge adds a serendipituous fourth. The Spanish working horse (Fig. 4) found the fruits abandoned by their extinguished ancestors and spread these two trees so thoroughly that today they are viewed by Mesoamerican societies as native and natural. And, in the case of *Enterolobium cyclocarpum*, cattle are surrogate horses (Janzen, 1982a).

However, as the motorbike and car replace the horse today, and as the cattle industry fades, these two trees are left as very visible living dead scattered across the former ranch lands, their abundant fruits rotting below the parent tree, the newly germinated seedlings killed by fungal pathogens nourished by the annually replenished seed crop, and the rare escaped seedling killed by herbicides, grass fires, and cosmetic cleansing.

III. WHEN IS A TREE NOT LIVING DEAD?

Earlier I noted that if each member of a tree species were to abruptly fall over dead the moment that it ceases to be a reproductive member of its population in its ecosystem, there would be far stronger alarm cries across the tropics about extinction rates and realities.

However, the isolated tree, left an adult in the open as the forest is mined away from around it (Fig. 2), is not necessarily or automatically a member of the living dead, or at least not necessarily at that moment. At least two circumstances may help to avoid this label. First, the pollinator community and the seed dispersal community for that tree may still be of a structure such that they confer sufficient amounts and patterns of their services and do so with the new reproductive phenology that will be expressed by the tree in its "new" habitat. And males do have fitness. A plant may never set a fruit or never have a surviving seedling from its seed crop, yet it still may be very much a member of the reproducing population (e.g., Aldrich and Hamrick, 1998). Plants contribute pollen "outward" as well as receive it from unseen members of the population. There may be some circumstances where this or that member of the pollinator guild will in fact carry pollen from that isolated tree back into the forest. At least potentially this may remove the living dead label.

Second, the new pattern of seed/seedling/sapling safe sites for that species may be sufficient for population survival, even if different. A novel demography, reproductive phenology, and microgeographic structure will ecologically emerge, reflecting the serendipituous matching of the tree's traits to these new conditions.

For the survivor, such ecological fitting (Janzen, 1985) of an individual (or a population) into the environment newly thrust upon it is the same process as occurs when a tree species is anthropogenically introduced to a new place. Whether introduced by humans or by natural processes, its survival there demonstrates that it has ecologically fit in. Such introduction may occur into a natural ecosystem or one variously anthropogenically perturbed. Sloppy deforestation may create many living dead, only mildly impact some other species, and favor yet new introductions into the region by having removed competitors or consumers.

A population of plants in a newly altered landscape is not necessarily at a given moment either "dead" or "alive." Just as the relationships of an individual to its ecological circumstances may decay slowly, it is also easy to visualize a population being sufficiently anthropogenically impacted that it gradually decays over several decades-to-centuries-long generations. This state of decay is an intermediate between living dead and "normal surviving." The portion of a population of trees at some geographic point may be in a constant state of swinging between being "okay" and living dead, as its associated climate and community of interactors goes through their own changes.

A species' population in its totality may also be waxing or waning in geographic coverage, density, "living deadness," or all three. Living dead are found at the geographic or demographic margins of all populations. It is just that human activity in ecosystem modification (elimination, simplification) simultaneously impacts so many species, and is so omnipresent, that it creates large numbers of living dead in the same place at the same time. These then carry the tragic perceptual load of tricking us into thinking that all is much more well than it actually is.

But ecological neutering, expressed as here in the terms "living dead" or "latent extinctions," is not restricted to the circumstance of the single tree in the field or a single portion of a population. The living dead are an integral part of natural age-structured mortality. Any field biologist can identify a large number of young individuals—seeds, seedlings, saplings—that have a vanishingly small chance of survival as individuals. The forest understory is densely populated with them, as is each squirrel's winter seed cache, as is the patch of seedlings below the healthy parent tree, as is the ground covered with ephiphyte seeds that fell past the branches of the trees above, as is the floor of the cave littered with bat-dispersed seeds. A very large part of the world's herbivore machine is run with this fuel and actually should be labeled "detritivore" rather than herbivore. The implications for evolutionary biology are huge, given that no matter how much herbivory occurs on these living dead, there can be no natural selection inflicted on the food populations.

Living dead adult individuals are also a prominent part of many undisturbed habitats and ecosystems. These are the waifs, the strays. Each of these is a plant whose seed arrived, grew to an adult, but found itself in a place lacking whatever is needed to maintain a viable population (Janzen, 1986c). In complex interwoven tropical habitats and ecosystems, the species list in a given place may contain as many as 10 to 20% of these kinds of living dead. For example, if a valley-bottom forest is eliminated, over time a significant number of tree species may disappear from the adjacent ridge, not because of any direct impact on the ridge forest but because the portions of the populations that were there are no longer maintained by seed flow into them from the valley bottom. This phenomenon is particularly visible where a particular soil or slope is thoroughly cleared for a crop, and the natural vegetation is left relatively intact in a neighboring habitat, ostensibly to protect it. Some species disappear because the conserved habitat did not really have its mutualist animals and physical climate conserved, or because it is too small, but others disappear simply because they were naturally occurring living dead.

Not to belabor the obvious, a tree standing dormant in the tropical dry season is not reproducing in the narrow sense, but it is also not necessarily living dead. But this is tricky for the observing human. We are very accustomed to being around trees that are not, at that moment, undergoing anything that appears to be reproduction, yet are members in good standing of quite surviving populations. The living dead tree does not display anything much different at first glance. Recognition of living dead status requires in-depth knowledge of its activities over decades, requires knowing if and where its pollen is going, and requires knowing where its seeds are moving to and what happens to them when they get there. This understanding is not acquired with the casual glance (e.g., Aldrich and Hamrick, 1998; Curran *et al.*, 1999; Hallwachs, 1986).

IV. WHAT OF SMALL PLANTS?

The isolated tree in the pasture has been a convenient illustrative example, but the world to which these ideas apply is far greater than that of large tropical trees. A small herbaceous plant may be a perennial with longevity like that of a tree. When the euglossine bees are extinguished through forest partial clearing, a *Catasetum* orchid they pollinated is left high on the main trunk of a shade tree left behind, a living dead in its own right. The orchid may flower for a century, waiting in vain for its long-distance pollinators (Janzen, 1974). They are long gone, their year-round nectar and pollen sources turned to charcoal. A living dead clump of perennial grass on a landslide scar may for many decades produce its small hard seeds, designed millions of years ago for a trip through a seed-dispersing, now-extinguished, large herbivore to a new disturbed site (Janzen, 1984). It finally succumbs to its individual sterile fate as the landslide scar revegetates to forest. A living dead herbaceous morning glory (Convolvulaceae), sprouting and flowering year after year into the insolated roadside ditch from its underground tuber, may never again see the bees that once moved among its flowers and the flowers of the many other forest-edge species that once sustained them (e.g., Frankie *et al.*, 1998).

But as mentioned earlier for a population of trees, even a population of annuals may also be a living dead population. Yes, each year it may flower and seed and disperse and then again germinate with the next rains. But did it make "enough" seeds? Were they set at the "right" time? Did they have the right genetic composition? Did they move to the right safe sites? Were those sites there to be moved to? Does the population do all this and much more to hold its place in the naturally shifting nature of its surroundings? Each year the population may decline a bit. Maybe even in some years it recovers. But overall, gradually it slides into local extinction.

Looking backward at the history of a plant population "going extinct," it may be possible to describe the decay of such a living dead population. Looking forward, however, it is much harder to label than is the living dead tree in a cornfield. After all, all populations have their ups and downs. How to know, other than retroactively, when a down is a downswing versus a slide into extinction? When the habitat destruction is major and obvious, the prediction is much easier, but perhaps more scientifically trivial, than when the habitat destruction is piecemeal, fuzzy, or widespread yet light.

V. WHAT OF ANIMALS, THOSE THINGS THAT MOVE?

Reproduction—that is, membership in the population—has two components. On the one hand, it is self-evident that the individual needs to be physiologically able to reproduce. On the other hand, if it is ecologically neutered, it is as dead as if sliced off with a chain saw. Selection has not generally favored the ability of a tree to "know" that it has been ecologically neutered by the removal of its pollinators, its dispersal agents, or the safe sites for its juveniles, and then take remedial action. What would the mutant tree have to be able to do? Walk back to the forest? Animals, with their chance to move to a new ecological circumstance, get horny. They search for nesting sites and mates, they may fight harder for their surviving fewer children, or they may migrate or emigrate to other places. But, in the face of the sweeping and omnipresent hand of humanity, busily extending its extended genome to cover the globe with both people and their domesticates (Janzen, 1998), where is the potentially living dead animal to go, and how long does it have to get there? One can search only so long before dying of old age, becoming a road kill, or running out of stored food reserves.

The tropical agroscape, and most wildlands as well, are awash with living dead animals, animal populations, and animal arrays (also known as "communities," whatever those are). Latent extinction is everywhere, but it operates more rapidly on animals with their high turnover rate and their lower capacity for extended lives as dormant seeds, resprouting root stocks, clonal patches, and so on.

Humans contribute in a curious perceptual manner to us being less aware of the animal living dead. At the level of the large animals, "everybody knows" that jaguars and tapirs are still "here" because everyone knows someone who knows someone who saw one once. One sighting of one 10-year-old jaguar crossing the road at noon 12 years ago will sustain the living dead jaguar in that area for decades, long past its consignment to the litter. It has taken more than three decades for the myth of Costa Rican giant anteaters, which once ranged these forests, to die a natural death.

Collectors and collections do their part as well. There is a snapshot of history present in our museum drawers, each specimen with its neat locality label. These collections continue the illusion of survival long past the reality. Retroactive data capture from museums gives a distribution map not of what is today on the Costa Rican countryside, but rather what once roamed where today sweeps unbroken waves of sugarcane, pasture, plantations, and horticulture. Intellectually every taxonomist knows this, but the orderly march of specimens across the museum drawers that read Panama, Costa Rica, Nicaragua, Guatemala, Veracruz, and San Louis Potosí lull one into thinking "surely over that huge geographic range there are still viable populations." Plants are not immune to these processes. It is just that with the more illusive, the shorter lived, the more mobile, the animal living dead may be more easily manifest in historical collections than on looking out the car window at 70 kmph.

And, when one descends from a field vehicle somewhere, a rare butterfly flutters from the museum drawer and down the roadside ditch, the cruel illusion is reinforced. Highly mobile animals are particularly effective at hiding the living dead from perception. The last living dead Costa Rican green macaws will fly across the countryside for decades. One small viable population of butterflies can create hundreds of living dead individuals searching across the food-plant-free agroscape until dying on windshields, of pesticides, or in the collector's net.

Some animals, like some plants, thrive in the agroscape. Are they living dead as well? The agroscape changes its biotic and its physical traits at the whim of some combination of the market and our technical ability to (re)engineer our domesticates (and produce new ones). Overnight the agroscape can flip from heaven to hell for a particular species. When cotton was the crop of choice on the Costa Rican countryside, the world was an ocean of food for native *Dysdercus* cotton-stainer bugs (as well as for a number of other native cotton herbivores). The local extinction of the bugs' original wild food plants (Malvaceae, Sterculiaceae, Bombacaceae) that accompanied the forest clearing for cotton fields was invisible. But when the downstream shrimp industry decided that it could no longer tolerate the pesticide runoff from the cotton fields, and cotton went the way of history, then so did the populations of cotton stainers. Some remain on as tiny (living dead?) populations on the seeds of local roadside malvaceous and sterculiaceous herbs, but even these may be living dead with their food plants easing their slide into extinction.

Does the ecologically neutered tree try harder, as an animal might? Could there be selection for such behavior? What does the isolated tree in the field perceive? What is perceived by an elephant-dispersed tree in a forest where the elephants have been extinguished? The tree in the field can know that much less pollen of this or that genetic composition now arrives, and may adjust accordingly—it may flower longer, it may set more seeds that are fertilized with its own pollen. It may make more flowers more regularly or it may set more wood or grow a larger crown. All of these things are simple responses to a circumstance that must occur in a natural forest to this or that individual that is not living dead. But the extinction of animal dispersal agents and safe sites for juvenile plants goes unheralded, with not even a potential feedback loop.

VI. AND WHAT OF THE THINGS THAT EAT THE LIVING DEAD?

All have their predators, their parasites, their mutualists, their scavengers. Many of these are quite dependent on the traits of their hosts. Food is not food is not food. Narrowly host-specific specialists abound.

For every living dead individual, population, or species, there is a large suite of consumers—individuals, and even species—living at the margin of their existence. A seed predator weevil—*Rhinochenus stigma*—passes its larval stages in the pods of guapinol (*Hymenaea courbaril*) on the Costa Rican countryside (Janzen, 1974). It maintains what appears to be a healthy population in the annual to supra-annual fruit crops that are destined to fall and rot below the parent in the absence of both the Pleistocene megafauna and the agouti (*Dasyprocta punctata*), contemporary inheritor of the guapinol (Hallwachs, 1986). But as each of those old guapinol trees dies at the end of its 200 to 500 year life span, the weevil population takes another hit. One day the last living dead guapinol trees will die, and along with them will go what appears today to be a perfectly healthy community of weevils.

The guapinol is also fed on by leaf-eating caterpillars. One, a large saturniid, *Schausiella santarosensis*, eats only guapinol leaves and will go the way of the Rhinochenus weevil. Another, *Dirphia avia*, also a large saturniid, feeds also on the foliage of Spanish cedar (*Cedrela odorata*), mahogany (*Swietenia macrophylla*), oak (*Quercus oleoides*), and guarea (*Guarea excelsa*) (Janzen and Hallwachs, 2000). As the adult guapinol trees dwindle in number, how the *Dirphia avia* population will twist and change will depend in part on how many individuals of the other living dead remain. (You guess: How many Spanish cedar, mahogany and oak trees will be left standing by the Costa Rican roadside?) Perhaps *Guarea excelsa*, its wood of no commercial value, will be the only host plant left. Enough to sustain *Dirphia avia*? Who knows, but it certainly won't be the same moth population that it was before.

The flowers of the living dead *Andira* trees were once a primary food source for tens of thousands of individuals of hundreds of species of bees; today they are visited by only a pale shadow of this bee community (Frankie *et al.*, 1998). But those old adult *Andira* con-

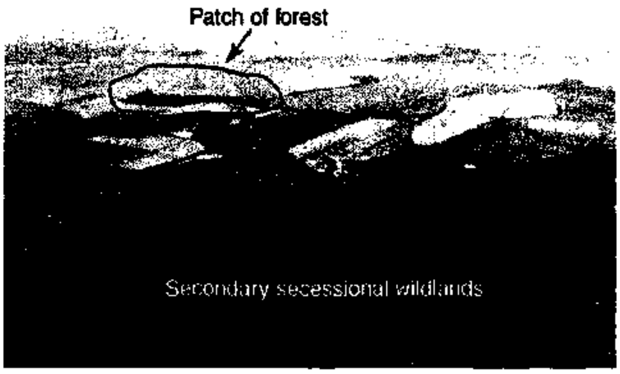


FIGURE 5 A living dead patch (left center) of natural vegetation, composed primarily of living dead individuals, among rice fields. There is essentially no gene flow between the patch and the secondary successional wildland in the foreground despite the thin connecting strip of riparian vegetation. Southwest of Liberia, Guanacaste Province, Costa Rica, December 14, 1999.

tinue to produce their massive flower crops and will do so for many decades to come. Its copious fruits, now largely from pollination by domestic honey bees, lie rotting below their parents in the absence of the masses of frugivorous bats that once dispersed them (Janzen *et al.*, 1976).

As noted earlier, the living dead are a "natural" part of any plant population. They are those individuals that have fallen where they have no chance of survival to reproduction. There are even living dead that have lived past their reproductive age. However, these living dead differ from the tree in the field in a very critical way for those who consume them. These living dead are being continually replenished by the natural dispersal process. They do not herald an invisible walk to extinction for the consumer.

VII. ARE THERE LIVING DEAD HABITATS AND ECOSYSTEMS?

Even when heavily agroindustrialized, the tropical agroscape often has patches of wildlands (Fig. 5)—forests along rivers and ravines, broken topography, swamps and marshes, vegetation on bad soil, no-man's land between rival owners, woodlots, hunting preserves, industrial accidents, parks, and parklets. This remaining natural vegetation is a patchwork and a dot map, and it appears to be 1 to 20% of the original vegetation. And it gives one hope.

One says, "aha, there are remnants. There is wild biodiversity on the countryside, in the agroscape. There is hope outside of the reserves" (which are so hard to maintain and seem so expensive in national park status). This is a cruel illusion. Descend to one of these patchlets of forest, so green, so tree-filled. It is a biodiversity desert, lacking 50 to 99% of its original biodiversity that it had when it was once part of a forested landscape. As a package it is a vegetational living dead. Its species list is a mix of actual living dead and a few species that can maintain viable populations under these circumstances. Our major problem is that we visit these patches as tourists. We were not there in 1965 to see their earlier biodiversity, to compare it with its pale shadow in 1999 (but see Frankie *et al.*, 1998).

Why are the survivors living dead, and what happened to those that have gone locally extinct? Part of them went when the area got so small that there were no longer circumstances for a viable population size. Part of them were explicitly mined or hunted. Part of them went when their mutualists, prey, and hosts went. Part of them went when the neighboring habitat, a habitat that spit seeds into the remaining forest and thereby maintained a population there, went to croplands. Part of them went when the seasons got drier, or wetter, or windier, or more fire-rich, or longer, or shorter, or, or, or.

Even those national parks that seem so secure are at major risk from this phenomenon. When the Southeast Asian dipterocarp trees fruit, the wild pigs come from everywhere and the collective seed crop of the preserved forest patch has no chance of satiating these seed predators (e.g., Curran *et al.*, 1999). It may be better to surround a conserved wildland with wild animal-free rice fields than oceans of secondary succession subsidizing waves of animals that then turn the small old-growth forest into yet more secondary succession by defecating seeds all over it (e.g., Janzen, 1983a).

The bottom line is that the complex fabric woven from thousands of interacting species has been ripped to bits. Many of those that seem to have survived are living dead, or the serendipituous few that find this new impoverished habitat to their competitive liking. In short, these patches are only pseudo-remnants, not really smaller pieces of what once was. Even those ecosystems and habitats that have always existed as small units—a marsh, a landslide scar, a volcano top, a patch of serpentine soil—did not live in isolation. Rather, each was maintained by a complex ebb and flow of immigrants, waifs, and influences from the neighbors. When the neighboring natural system is turned to cropland, the integrity of the small natural patch (e.g., Fig.

5) is usual trashed almost as badly as if an army of chain saws had run through it. It just takes a bit longer for the living dead to live out their physiological lives.

These impoverished patches are especially deceptive for the bioilliterate. For those to whom a forest is just a batch of large woody plants, for those who cannot or will not read the differences between an advertising ditty and a complex poem, the agroscape with its living dead and pseudo-remnant natural vegetation appears to be not much different from a glade and forest mix in a national park. All seems to be well. But when humanity expects something from that wildland patch, it discovers that almost all of its tropical biodiversity is gone.

These patches have also played a mean trick on the conservation community. A huge portion of the world's conservation policy is based on the understandings of nature held largely intuitively by those who have grown up extra-tropical and learned their lessons from extra-tropical ecosystems. They easily adopt the mantra of trying to save the biodiversity remnants scattered across the agroscape. They are especially prone to do so in the face of the frustration of trying to save very large (and commercially juicy) blocks of intact vegetation. The forest-patchlet-dotted agroscape of Minnesota or Sweden still collectively contains easily more than 80% of the species that were there when the European colonists arrived. However, the same snapshot of a Costa Rican agroscape contains at best 5 to 20% of what once was. And the percent is still falling rapidly because a huge fraction of what remains today is living dead.

The more biodiverse and the more complex an ecosystem, the more likely that human perturbation will create anthropogenic living dead among the species with longer-lived individuals. This is because perturbations strip away mutualists and other biointeractors, leaving behind the physiologically functional individuals to live out their neutered life spans. The more biodiverse and the more complex, the more likely any given individual is to be dependent on one or more of these interactants to remain a member of the population.

This tropical-to-extra-tropical comparison, derived by spending my life peering closely at both tropical and extra-tropical habitats is a major driver behind the conclusion that in the tropics a triage decision is needed. The living dead are writhing in lethal pain on the battlefield of the tropical agroscape. If we expend our scarce financial, political, and social resources on them instead of saving a few large coherent blocks of multi-ecosystem biophysical units, in the end we will live an even yet more impoverished biodiversity existence.

The future of real conservation in the tropics lies in by-and-large focusing our efforts on the survival of a relatively small number of very large and diverse biophysical units, each complicatedly integrated with local, national, and international societies (Janzen, 1998, 1999). Painful as it may be, resources spent on trying to save individual species and small habitat fragments scattered across the agroscape, often living dead, is bad conservation economics and creates an angry antagonistic *Homo sapiens*.

We have no option in the tropics but to recognize that conserved wildlands are and always will be islands in an ocean of agroscape. Our task is to get on with rendering them into the highest quality islands possible, and not be distracted by, nor lulled by, the living dead individuals and islandlets. Yes, if there remains but just one Rembrandt painting, we of course save it even if it is bullet-holed and faded. However, we must recognize it for what it is and not convince ourselves that by doing so we have preserved our knowledge of European history.

VIII. RESTORATION BIOLOGY

The living dead are largely a negative force in the algebra of conservation biology and conservation reality. However, in those few cases where ecosystem restoration is desired or serendipitous, their life span delimits a window of opportunity for the reintegration of their species into the restoring ecosystem. Reintegration is not an unqualified given, however. A single large tree in a pasture being restored to forest may be dropping its seeds and fruits into an early successional old-field community that for decades is still way too unattractive to contain the seed dispersal coterie that will begin to restore the demography of that tree species. Equally, the pollinators of its flowers may already be extinct, or abhor the young secondary succession coming up below the large old parent. And finally, the physical climate of the highly deciduous and dry-season blasted secondary succession may well be a dismal place for a seedling or sapling of that old-growth giant. As every plantation initiator knows, the act of stuffing seeds into the ground does not a plantation make.

Until a very short time ago, the California condor was made up of living dead individuals. They were brought into captivity (e.g., transplanted to a safe field), reproduced (e.g., seeds collected and grown in pots), and have been put back out, hopefully in an agroecosystem with a friendly sociology. This habitat is, however, very seriously impoverished through reduction of marine mammal populations that so kindly generated the cadavers for lunch, and the California condor may always be dependent on human subsidy.

Many species of living dead may be rescued in this manner, if we care enough to spend the resources on them and gather information about them. But before racing out to apply the same technique to the living dead guapinol trees in the centers of Costa Rican pastures, a question very much needs to be addressed. Would not the same money spent on saving large blocks of guapinol-occupied wildlands, complete with their pollinators and dispersal agents, not generate vastly more conservation of guapinol and its hundreds of thousands of compatriot species? Yes, even these large blocks of wildland will contain some living dead. The wildland's biodiversity will attain an equilibrium density at whatever number of species survive the reduction from a continent of wildland to a large island of wildland. Those who are extinguished during this process will suggest the list of who were the living dead.

See Also the Following Articles

CENTRAL AMERICA, ECOSYSTEMS OF • CONSERVATION BIOLOGY, DISCIPLINE OF • DEFORESTATION AND LAND CLEARING • EXTINCTIONS, MODERN EXAMPLES OF • FOREST ECOLOGY • MAMMALS, LATE QUATENARY, EXTINCTIONS OF • POLLINATORS, ROLE OF • RANGE ECOLOGY • RESTORATION OF BIODIVERSITY • TROPICAL ECOSYSTEMS

Bibliography

Aldrich, P. R., and Hamrick, J. L. (1998). Reproductive dominance of pasture trees in a fragmented tropical forest mosaic. *Science* 281, 103–105.

Curran, L. M., Caniago, I., Paoli, G. D., Astianti, D., Kusneti, M., Leighton, M., Nirarita, C. E., and Haeruman, H. (1999). Impact of El Niño and logging on canopy tree recruitment in Borneo. *Science* 286, 2184–2188.

Frankie, G. W., Vinson, S. B., Rizzardi, M. A., Griswold, T. L., O'Keefe, S., and Snelling, R. R. (1998). Diversity and abundance of bees visiting a mass flowering tree species in disturbed seasonal dry forest, Costa Rica. *Journal of the Kansas Entomological Society* 70, 281–296.

Hallwachs, W. (1986). Agoutis (*Dasyprocta punctata*): The inheritors of guapinol (Hymenaea courbaril: Leguminosae). In *Frugivores and Seed Dispersal* (A. Estrada and T. Fleming, Eds.), pp. 285–304. Dr. W. Junk Publishers, Dordrecht.

Janzen, D. H. (1974). The deflowering of Central America. *Natural History* 83, 48–53.

Janzen, D. H. (1981). *Enterolobium cyclocarpum* seed passage rate and survival in horses, Costa Rican Pleistocene seed dispersal agents. *Ecology* 62, 593–601.

Janzen, D. H. (1982a). Differential seed survival and passage rates in cows and horses, surrogate Pleistocene dispersal agents. *Oikos* 38, 150–156.

Janzen, D. H. (1982b). How and why horses open *Crescentia alata* fruits. *Biotropica* 14, 149–152.

Janzen, D. H. (1983a). No park is an island: increase in interference from outside as park size decreases. *Oikos* 41, 402–410.

Janzen, D. H. (1983b). The Pleistocene hunters had help. *American Naturalist* 121, 598–599.

Janzen, D. H. (1984). Dispersal of small seeds by big herbivores: Foliage is the fruit. *American Naturalist* 123, 338–353.

Janzen, D. H. (1985). On ecological fitting. *Oikos* 45, 308–310.

Janzen, D. H. (1986a). The eternal external threat. In *Conservation Biology: The Science of Scarcity and Diversity* (M. E. Soule, ed.), pp. 286–303. Sinauer Associates, Sunderland, MA.

Janzen, D. H. (1986b). The future of tropical ecology. *Annual Review of Ecology and Systematics* 17, 305–324.

Janzen, D. H. (1986c). Lost plants. *Oikos* 46, 129–131.

Janzen, D. H. (1998). Gardenification of wildland nature and the human footprint. *Science* 279, 1312–1313.

Janzen, D. H. (1999). Gardenification of tropical conserved wildlands: Multitasking, multicropping, and multiusers. *PNAS* 96(11), 5987–5994.

Janzen, D. H., and Hallwachs, W. (2000). Philosophy, navigation and use of a dynamic database ("ACG Caterpillars SRNP") for an inventory of the macrocaterpillar fauna, and its food plants and parasitoids, of the Area de Conservacion Guanacaste (ACG), northwestern Costa Rica (http://janzen.sas.upenn.edu).

Janzen, D. H., and Martin, P. S. (1982). Neotropical anachronisms: The fruits the gomphotheres ate. *Science* 215, 19–27.

Janzen, D. H., Miller, G. A., Hackforth-Jones, J., Pond, C. M., Hooper, K., and Janos, D. P. (1976). Two Costa Rican bat-generated seed shadows of *Andira inermis* (Leguminosae). *Ecology* 56, 1068–1075.

LATITUDE, COMMON TRENDS WITHIN

Michael R. Willig
Texas Tech University

GLOSSARY

alpha diversity The diversity of species, often estimated as species richness, within a local community or site.

beta diversity The degree of turnover in species (and changes in their abundances) among communities or sites along a gradient or within a larger area.

gamma diversity The diversity of species, often estimated as species richness, in a larger area as a consequence of both alpha and beta diversity.

latitudinal gradient A gradual change in a characteristic of interest (e.g., species richness) with a gradual change in latitude; a gradient is well defined if it adheres to a particular mathematical relationship.

middomain effect A gradient wherein species richness increases symmetrically from latitudinal extremes to the middle of a region as a consequence of the random placement of species ranges within a geographic domain (also known as Perinet effect).

rapoport effect A latitudinal gradient wherein the sizes of the distributional ranges of species decrease with decreasing latitude.

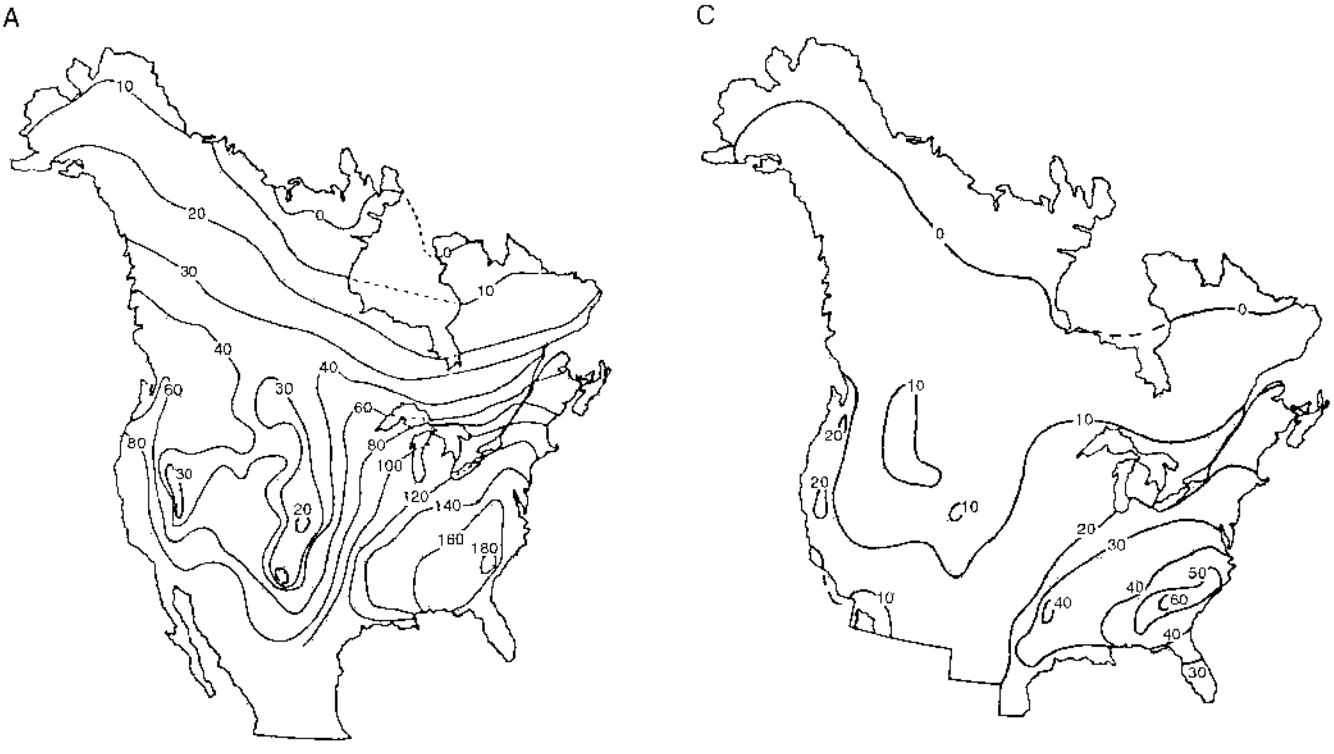
scale dependence A condition in which either the form or the parameters of a relationship between two variables (e.g., richness and latitude) is contingent on spatial or temporal attributes.

species density The number of species within a sampling unit of fixed size.

species diversity A feature of biological communities or assemblages that reflects the variety of organisms in an area and that includes two components, species richness and species evenness (the degree to which all species have the same proportional abundance).

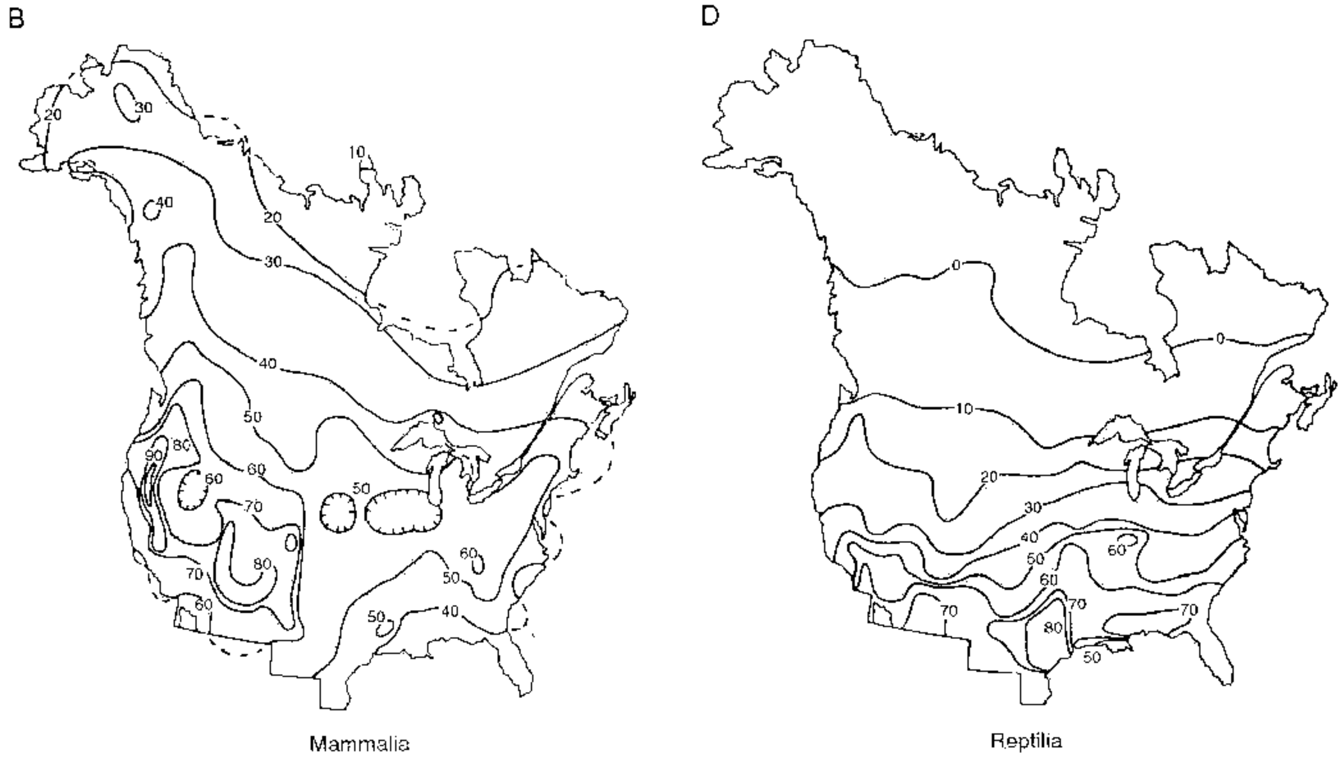
species richness The total number of species in an area.

LATITUDINAL gradients of diversity are biogeographic patterns that define the way in which the number of species changes with latitudinal position on the surface of the earth. The general pattern is for species richness to increase from polar to tropical regions (Brown, 1995; Gaston, 1996; Rosenzweig, 1995), regardless of the taxonomic affiliation of the organisms (e.g., mammals, fishes, insects, and plants) or geographic setting in which they occur (e.g., Africa, South America, and the Atlantic Ocean). This is true for extant organisms (Fig. 1) as well as for those organisms alive during the past 70 million years (Fig. 2). An increase in species richness with decreasing latitude is the pattern generally observed at three spatial scales, including the level of



Trees

Amphibia



Mammalia

Reptilia

FIGURE 1 Species richness gradients in Canada and the United States for trees (A), mammals (B), amphibians (C), and reptiles (D). Contour lines connect localities with approximately equal species richness (reproduced from Currie, 1991, with permission from The University of Chicago Press).

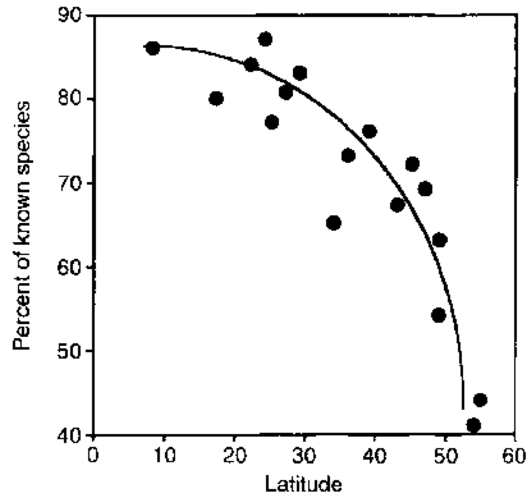


FIGURE 2 Latitudinal gradient in fossil species richness for marine Foraminifera from approximately 70 million years ago (modified from Rosenzweig, 1995).

broad climatic zones (Fig. 3), assemblages occupying arbitrary geographic subdivisions (i.e., quadrats or bands) of the earth's surface (Fig. 4), and local ecological communities (Figs. 4 and 5). Nonetheless, not all taxa increase with latitude in the same fashion, and a few groups do not even exhibit the general pattern of a latitudinal increase in richness. Moreover, considerable controversy surrounds the mechanisms that affect latitudinal patterns in diversity, with ecological, evolutionary, historical, and stochastic processes (Table I) championed as the cause(s) of observed gradients (Rhode, 1992). Indeed, many of the mechanisms are circular or unsubstantiated by empirical data.

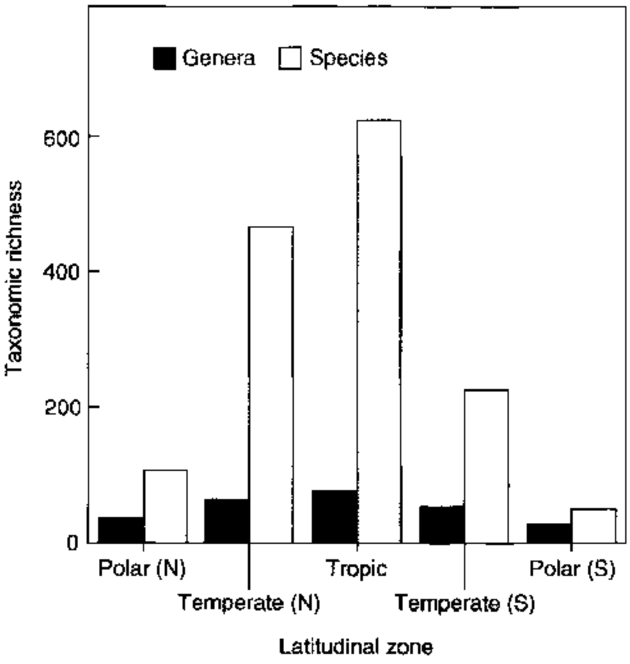


FIGURE 3 Generic and species richness of tunicates in each of five major climatic regions defined by latitude and arranged from northern-most latitudes, through the tropics, to southern-most latitudes (modified from Fischer, 1960).

I. CONTEXT

Since the voyages of Darwin and Wallace, biologists have been fascinated with the high species diversity of tropical regions compared to those in temperate or boreal zones. Indeed, this fascination with tropical diversity catalyzed in many ways the conceptual development of the theory that currently constitutes modern ecology. Moreover, increasing concern about the loss of diversity, especially in tropical regions, has led to the rapid development of the science of conservation biology. Documenting the way in which diversity differs across the globe and understanding the mechanisms that produce such variation are critical steps in the design of global conservation strategies and the implementation of regional management plans.

The mid-1950s to early 1960s saw the emergence of rigorous quantification of broad-scale relationships between species richness and latitude (Fischer, 1960). Within the next 25 years, scientists convincingly had documented the ubiquity of gradients in which species richness increased toward tropical areas. Similar gradients also were documented for diversity of higher taxonomic groups (e.g., genera, families, and orders) (Figs. 3 and 4). Indeed, the increase in species richness for terrestrial and marine environments was quantified successfully for a wide variety of taxonomic groups, such as mammals, birds, reptiles, amphibians, fish, tunicates, crustaceans, mollusks, brachiopods, corals, foraminiferans, and vascular plants. Nonetheless, some taxa representing lower levels in the systematic hierarchy (i.e., orders or families) were notable exceptions in having maximal diversity in polar (e.g., seals, penguins, and sandpipers) or temperate zones (e.g., voles, salamanders, ichneumonid wasps, and coniferous trees). Caution must be employed in considering such exceptions because other groups of equivalent rank within the same higher taxon often are restricted to lower latitudes, and the higher taxon exhibits a tropical maximum in species richness.

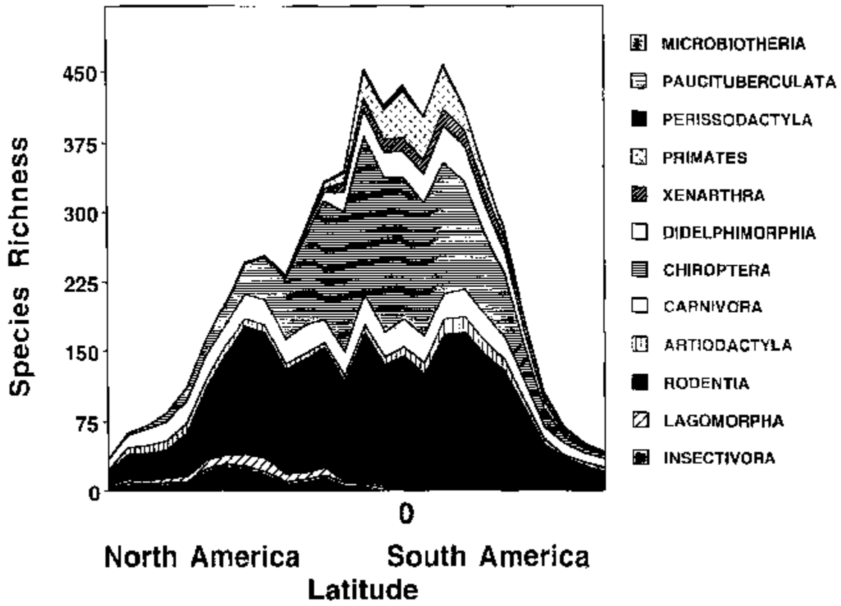


FIGURE 4 Differential contribution of mammalian orders to the latitudinal gradient of species richness in the New World based on data from 5-degree latitudinal bands (after Kaufman, 1995).

II. PATTERNS

A gradient implies a gradual change in species richness with a gradual change in latitude. In an unambiguous fashion, the form of that pattern is the precise mathematical or statistical relation that describes how species richness changes with latitude. As a consequence, three considerations are important in assessing patterns: the general shape of the curve (e.g., symmetry, kurtosis, or linearity), the parameters that characterize the relation, and the degree to which the fit of empirical data to the predicted curve is equivalent to the north and south of the equator. Knowledge of these three aspects of gradients suggests the kinds of causal mechanisms that are in operation. In addition, it facilitates comparison of gradients among taxa within the same geographic domain (Fig. 1; birds versus mammals versus reptiles versus amphibian in North America) as well as comparisons among different geographic domains for the same taxon (Fig. 4; North American versus South American for mammalian orders).

Patterns are often scale dependent, with particular mechanisms more likely operating at some areal scales than at others. Consequently, patterns will be elucidated for each of two foci: biotic assemblages occupying broad areas and ecological communities occupying local sites. These scales are intimately associated with each other. In part, the diversity of regions, biomes, or climatic zones is a consequence of the species richness that is accumulated within local communities. Similarly, the species richness and composition of local communities are affected by the set of taxa that constitute regional species pools (Putman, 1994).

A. Assemblages

Most of the empirical research concerning the relationship between species richness and latitude that has been done using arbitrary sampling units has been based on (i) latitudinal bands, (ii) quadrats of fixed area, or (iii) quadrats of unequal area defined by lines of longitude (meridians) and latitude (parallels). Alternatively, research has focused on the species richness of biomes or broad latitudinally defined climatic zones. Because the area of any sampling unit may have as large or larger an effect on variation in species richness than does its latitude, it is critical to understand how area may affect latitudinal patterns in different ways, depending on the method or approach.

Generally, analyses of quadrats defined by meridians and parallels are inferior to those based on other sampling units because such quadrats differ in area in a systematic fashion and bias quantitative conclusions. As meridians converge toward the poles, the size of the

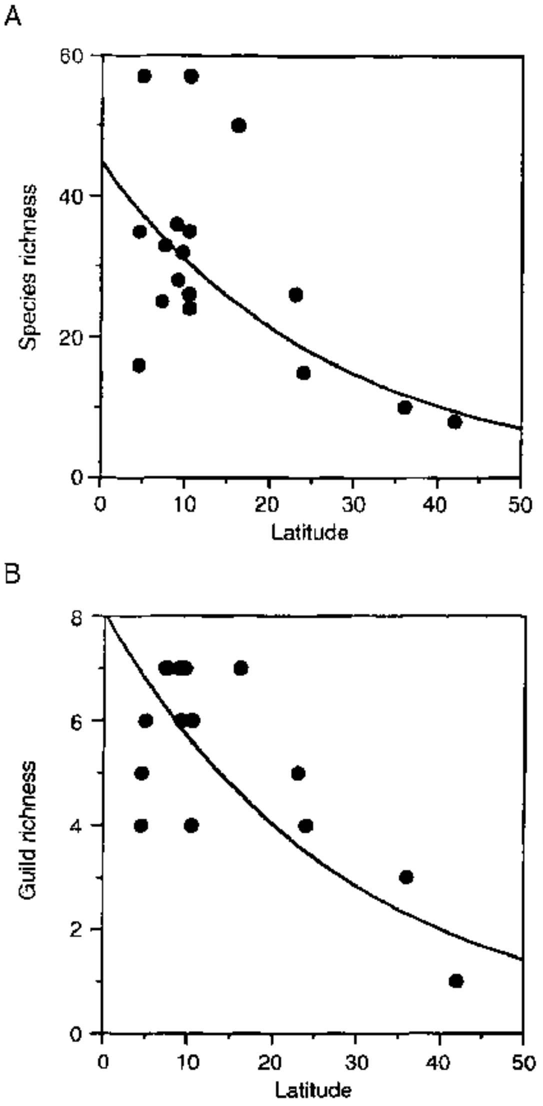


FIGURE 5 Latitudinal gradient of species richness (A) and feeding guild richness (B) within local communities for bats from throughout the New World (based on Stevens and Willig, 1999).

quadrats becomes smaller. Consequently, any attempt to control for variation in richness among quadrats as a consequence of area will remove at least some of the effects of latitude as well. This confounding effect compromises the detection of pattern across broad latitudinal gradients.

Analyses based on latitudinal bands also must control for the effect of area because the width of a continent is not constant at all latitudes. A variety of methods have been used to compensate for this problem. Importantly, care must be employed when the area of bands varies in a systematic fashion with latitude because of the shape of the continent (e.g., progressive decreases in area with decreasing latitude in North America versus progressive increases in area with decreasing latitude in South America). In such cases, analyses based on bands may be plagued with the same confounding effects as those for analyses based on quadrats defined by meridians and parallels. Indeed, if classical statistical techniques are used to control for the effect of area in bands defined by 5-degree meridians in North America, the areal relationship is contrary to both common sense and ecological theory in that species richness increases as area decreases. Subsequent assessment of latitudinal effects may be severely compromised because the width of North America decreases as latitude decreases. Hence, adjusting the latitudinal gradient to account for area also removes an appreciable latitudinal effect. Nonetheless, if continental shape does not confound the effects of latitude and area, then regression techniques hold great promise, especially if appropriate nonlinear approaches are taken to adjust species richness in accord with species area theory (i.e., adjust richness of bands to a common area based on nonlinear regression of nested quadrats within each band).

Quadrats of fixed size also have been used to analyze broad-scale patterns of diversity. Nonetheless, variation in species richness among sampling units could still be a consequence of area, at least partially, when quadrats occupy coastal positions along continental borders. Adjusting for area in these cases may obscure the effects of rapid transition zones in terrestrial communities as they approach land–sea margins. Hence, the consensus is to not consider quadrats unless they are full of land. Subsequent variation among quadrats that is due to latitude can be assessed through a variety of statistical models. However, here too it is important to note that patterns are scale dependent. That is, the pattern detected for quadrats encompassing 100 km^2 could be quite different from those at 10,000 km^2. The importance of scale dependence in ecology and biogeography has been emphasized increasingly during the past decade. Nonetheless, broad-scale biogeographic patterns for sampling units between 1000 and 25,000 km^2 have been shown recently to be scale independent and little affected by area (Lyons and Willig, 1999).

Much of the early literature on latitudinal gradients in diversity was based on the species richness of broadly defined climatic zones (e.g., north polar, north temperate, tropical, south temperate, and south polar) or geopolitical units (e.g., countries, states, and provinces). Taxonomic richness was documented to increase from polar to tropical regions (Fig. 3). Even when values for richness were not adjusted for the areal extent of geopolitical regions (e.g., snake species in Argentina,

TABLE I

Mechanisms Potentially Affecting the Latitudinal Gradient in Species Richness[a]

Circular	Empirically unsubstantiated
Competition	Environmental stability
Mutualism	Environmental predictability (contingency)
Predation	Productivity
Epidemics	Abiotic rarefaction
Biotic spatial heterogeneity	Physical heterogeneity
Population size	Angle of the sun above the horizon
Niche width	Area
Population growth rate	Aridity
Patchiness	Seasonality
Epiphyte load	Number of habitats
Host diversity	Rapoport's rule (range size gradient)
Harshness	Ecological time
	Evolutionary time
	Temperature depedence of chemical reactions
	Solar energy
	Evolutionary speed
	Stochastic placement of species ranges

[a] Modified from Rohde (1992).

ant species the Americas, and breeding bird species the Americas), the polar to tropical gradient was obvious. Nonetheless, controversy currently surrounds the interpretation of such data when the focus is on broad climatic zones associated with latitude. Some suggest that the increase in diversity with decreasing latitude primarily is a result of more tropical regions having larger areas than their extratropical counterparts. In contrast, others suggest that extratropical areas are often larger but have fewer species than their more tropical counterparts, suggesting that latitudinal correlates other than area are the driving forces behind the gradient.

B. Ecological Communities

The way in which latitudinal variation in diversity at broad spatial scales (gamma diversity) is related to patterns of species richness at the level of local communities (alpha diversity) is unclear. In part, this is because the geographic boundaries of a community are difficult to designate and are ultimately arbitrary decisions. Moreover, until recently, little was known about the composition of local communities in tropical regions, making assessment of broad-scale latitudinal patterns of community richness a premature endeavor. Finally, it is unlikely that a single research scientist could gather sufficient data across many sites to assess latitudinal gradients in community richness with sufficient power to distinguish pattern from noise. Hence, compositional data must be compiled from the work of many different individuals, who often use different methods, designs, and sampling intensities, to quantify the gradient in a meaningful way. The concern expressed by Voss and Emmons (1996) regarding adequate sampling within a community must be tempered by the realization that community composition has a temporal dynamic. Communities represent suites of populations with the potential to interact, and thus they must be constrained to some extent by both time and space. Sampling regimes which extend over protracted periods of time (e.g., decades or longer) may have inflated estimates of richness and inaccurate assessments of species composition. Nonetheless, recent progress in this regard allows quantitative evaluation of patterns in a rigorous way, at least for some taxonomic groups.

When care is taken so that a local community is delimited as a geographic area in which constituent species have a high likelihood of interaction, it becomes clear that species richness increases from polar through temperate to tropical regions. The gradient can arise from an increase in the number of functional groups (e.g., guilds or ensembles) within a community as well as from an increase in the number of species per functional group. For example, data for volant mammals (bats) from 17 local communities (Stevens and Willig, 1999) that met rigorous standards for inclusion in analyses (i.e., well-delimited local areas must have been sampled intensively for at least 1 year to include all seasons in which species are active) clearly exhibit a latitudinal gradient of increasing richness with decreasing latitude (Fig. 5A). A similar increase in the number of functional groups or feeding guilds within a community (Fig. 5B) contributes to this trend. Specifically, north temperate bat communities are depauperate with 10 or fewer species, all of which are members of carnivorous feeding guilds (e.g., aerial insectivore, mollosid insectivores, or gleaning animalivores). In contrast, tropical and subtropical communities on average contained 33.7 and 20.5 species, respectively. Moreover, functional diversity of communites in these two regions was high (equaling 5.9 and 4.5 feeding guilds, respectively) and included sanguinivorous, piscivorous, frugivorous, and nectarivorous taxa in addition to the guilds represented in the temperate communities.

Species richness of nonvolant taxa (tetrapods) in mammalian communities from North America also shows a significant increase from polar to tropical regions (Fig. 6; Kaufman, 1998). Tropical communities contained approximately 45.5 species on average, whereas extratropical communities contained only approximately 25.8 species. Within these two latitudinally defined climatic regions, the relationship between richness and latitude was nonsignificant. Most important, an examination of species turnover for these communities revealed that beta diversity increased toward the tropics as well. Hence, increasing diversity of broad geographic or climatically defined regions (gamma diversity) with decreasing latitude may be a consequence of the increase in local diversity (alpha diversity) as well as the increase in differentiation among local communities within latitudinal regions. This appears to be the case for nonvolant mammals in the New World, at least at some spatial scales.

Many factors (e.g., productivity, competition, predation, and disturbance) have been suggested as the dominant forces affecting the composition and structure of local communities. Early theoretical and empirical work stressed the role of deterministic factors such as competition in molding community attributes. Subsequent focus on the distinction between equilibrial and nonequilibrial communities cast doubt on the universality of deterministic mechanisms in general and competition in particular and raised serious questions about the degree to which local communities were saturated by species. Indeed, variation in the degree to which local communities attain equilibrial richness may contribute strongly to the latitudinal gradient in diversity (Putman, 1994). The latitudinal progression from polar to tropical regions may represent a gradient in the degree to which stochastic density-independent mechanisms or biotic interactions dominate the forces affecting the abundance of local populations and the composition of local communities. Specifically, the species richness of a community may be a consequence of the severity, variability, and predictability of local environmental conditions. Low predictability and harsh conditions predispose communities to be regulated by abiotic parameters and to have low diversity. High predictability and conditions that are clement favor high diversity. To the extent that high solar insolation and warm temperatures represent favorable conditions, and low intra-annual variation in temperature and rainfall represent predictable conditions, tropical communities should be more species rich than their extratropical counterparts. In essence, the factors that affect elevated richness at the local scale likely contribute to enhanced gamma diversity of regions as well (see Section IV).

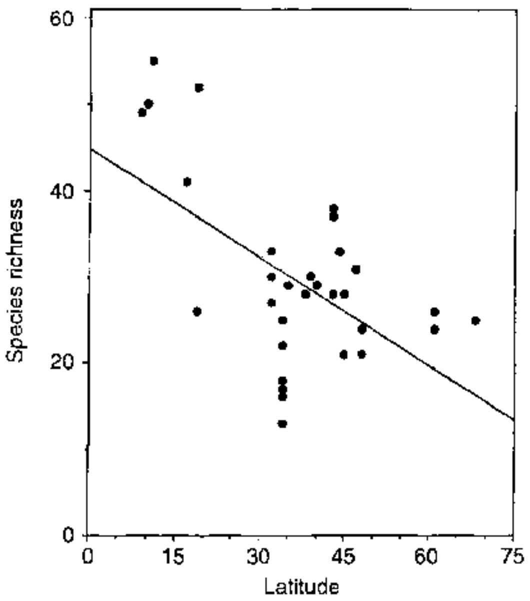


FIGURE 6 Latitudinal gradient of species richness for mammalian tetrapods in local communities in North America (based on Kaufman, 1998).

III. MECHANISMS

A grand proliferation of hypotheses (Table I), along with subsequent a posteriori modifications, is a characteristic feature of the literature concerning the relationship between diversity and latitude. A variety of factors contribute to this. Each of these hypotheses represents a conceptual model with only qualitative predictions. As with much of macroecological research, broad-scale data concerning the distribution of species is not available for many taxa. Manipulative experiments designed to disentangle the effects of competing hypotheses are not feasible or ethical. In addition, the inherent geographic factors that might affect richness are often correlated so that efforts to remove the effect of one to assess the other can lead to spurious results due to the confounded nature of the mechanisms.

Indeed, almost all hypotheses provide only a general qualitative prediction that richness should increase toward the tropics. The hypotheses represent conceptual models that lend insight into how nature could operate but do not generate unique predictions based on direct features of the gradient that allow conclusive elimination of competing hypotheses in the sense of strong inference. Moreover, many hypotheses are circular in nature and the indirect predictions that they make about latitudinal gradients have not been examined from an empirical perspective. In addition, some hypotheses only pertain to specific taxa or ecological groups so that they are not applicable universally.

Rather than elucidate the score of extant hypotheses, which has been done to greater or lesser extents elsewhere (Rohde, 1992; Rosenzweig, 1995; Gaston, 1996), an exposition of selected hypotheses which have generated debate in the contemporary literature follows. These hypotheses represent areas of research which likely will contribute to the complexion of the discipline in the future.

A. Geographic Area Hypothesis

The latitudinal gradient in which richness peaks in the tropics may be a consequence of the larger landmass of the tropics compared to other geographic zones. This simple idea had its genesis in the work of Terborgh (1973), with considerable development and refinement by Rosenzweig (1995) in subsequent years, during which the effects of productivity and zonal bleeding have been incorporated into a more comprehensive conceptual model. Nonetheless, the geographic area hypothesis has generated considerable controversy and contention (Rohde, 1997, 1998; Rosenzweig and Sandlin, 1997). The controversy does not surround whether an areal mechanism operates; rather, it focuses on the degree to which variation in area is the dominant factor molding latitudinal gradients in richness.

Two features of the earth's geometry predispose the sizes of tropical regions to be greater than those of their higher latitude counterparts. First, the earth is essentially a sphere. The distance between longitudinal meridians at the equator is greater than that elsewhere on the globe, and intermeridian distance decreases in a regular fashion toward the poles. Second, northern and southern tropical zones are adjacent, whereas the northern and southern variants of other latitudinally defined zones are isolated from each other. Nonetheless, the positions, sizes, and configurations of the earth's continents will affect the proportion of land or water in tropical versus extratropical regions, and this has varied over geological time as a consequence of plate tectonics. In addition, the number and breadth of zones used to subdivide latitude will affect the perception of areal dominance associated with the tropics. For example, a tripartite division (torrid, temperate, and polar) reveals that the tropics ranks second in area to north temperate regions at the global scale, with considerable variation in the proportional area represented by the tropics among continents (approximately 38, 12, 80, 41, and 0% of America, Eurasia, Africa, Australia, and Antarctica, respectively). In contrast, finer resolution of zones to tropical, subtropical, temperate, boreal, and tundra indicates the areal predominance of tropical lands globally.

Most important, the degree of environmental variation within the tropics is less than that in other geographic zones, at least with respect to incident solar radiation and temperature. Specifically, a band of 50° centered on the equator evinces no or little change in mean annual temperature with latitude (constant at approximately 27°C), whereas mean annual temperature decreases thereafter by approximately 0.75°C per degree latitude. Hence, regardless of the size of zonal subdivisions chosen to define tropical or extratropical regions, tropical landmasses are larger than any other landmasses with similar variation in temperature.

As a consequence of the areal extent and homogeneity of temperature and solar insolation in the tropics, speciation rates there should be higher and extinction rates lower than in extratropical regions. Specifically, the larger area of the tropics allows its species to have larger ranges than do their extratropical counterparts. Larger ranges allow species to be represented by more populations or populations of larger size, both diminishing the likelihood of extinction resulting from accidental cause or from environmental perturbation. Simultaneously, larger areas are more likely to contain or experience geological events that produce geographic barriers that enhance the rate of allopathic speciation. The dynamic balance between the rates of speciation and extinction therefore yield higher equilibrial richness in the tropics than in extratropical areas (Fig. 7).

Rosenzweig and colleagues marshal many lines of evidence in support of the geographic area hypothesis by citing two kinds of observations. First, larger biotic provinces, regardless of latitude, have more taxa than do their smaller counterparts (e.g., generic, familial, and ordinal richness of mammals increases with provincial area). Second, diversities from the same biome but from different continents or provinces differ as a function of their areal extent (e.g., rain forest vertebrates and plants as well as tropical freshwater fish increase in richness as their areal extent increases). Situations in which the general pattern does not occur usually include large but unproductive climatic zones with few species—effectively the richness-diminishing effects of low productivity may countermand the dominant role of area in these systems. In contrast, Rohde (1997) considers area not to be the dominant factor that affects high species richness in the tropics. He illustrates the point with Eurasian freshwater fishes, and shows that much smaller tropical regions have much greater species richness than do larger cold-temperate regions. Similarly, the expansive deep-sea biome with more or less constant temperature has far fewer species than its smaller tropical counterparts. Clearly, consensus is elusive concerning the relative importance of area in affecting species richness compared to other mechanisms.

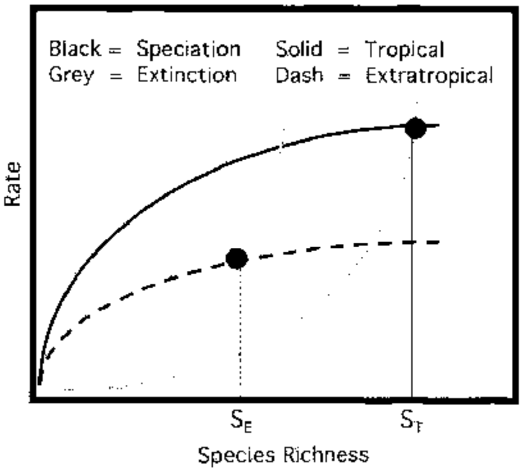


FIGURE 7 Graphical model illustrating the dynamic equilibrium between rates of speciation and extinction for tropical and extratropical latitudes (modified from Rosenzweig, 1995).

B. Evolutionary Speed

After a broad and incisive review of the various mechanisms purported to cause latitudinal gradients in species richness, Rohde (1992) found them all to be lacking. Instead, he suggested that the gradient was a consequence of differential rates of speciation associated with an important latitudinal correlate, temperature, rather than being a product of equilibrium-based ecological processes that presupposed that local communities are saturated with species. His conceptual model is erected on the foundation of three premises. First, tropical environments support shorter generation times for many homiotherms and poikiliotherms. Second, mutation rates increase as temperature increases and are highest in the tropics. Third, faster physiological processes occur at higher temperatures; this, coupled with the first two relationships, suggests an accelerated rate of fixation of favorable alleles in tropical populations. This effectively results in greater evolutionary time in the tropics for mechanisms of diversification to attain fruition.

Much of the subsequent dialog concerning the evolutionary speed hypothesis has been embedded in the debate concerning the efficacy of area, productivity, and zonal bleeding in producing empirical gradients of diversity with respect to latitude. Although this debate has been fruitful in crystallizing assumptions of equilibrial conditions and species saturation as they relate to the area hypothesis, it has not furthered our understanding of the possible role of temperature in effecting higher speciation rates. Rhode has challenged the research community with this supposition; unfortunately, the response is deafening in its silence.

C. Rapoport–Rescue Hypothesis

As its name implies, this hypothesis is a hybrid of two mechanisms operating in tandem: the Rapoport effect and the rescue effect. A geographic pattern in which species range size decreases from high to low latitudes recently has come to the forefront of the macroecological literature (Stevens, 1989, 1992) and been termed Rapoport's rule after the Argentine scientist who first discussed the pattern in the context of many other areographic principles. Stevens hypothesized that the latitudinal propensity for range size to decrease toward the tropics, when combined with differential movement of individuals from source to sink habitats (rescue or mass effect), can generate the latitudinal gradient of diversity.

Specifically, at any one locale in the temperate zone, an individual must be able to tolerate considerable intraannual variation in climatic conditions; thus, species that occur in the temperate zone can attain a wide latitudinal distribution because of the broad tolerance of its constituent individuals to varying local conditions. In contrast, an individual in the tropics experiences little seasonal variation in climatic conditions; consequently, species comprising individuals that occur in tropical zones are predisposed to have narrower latitudinal distributions. This creates the Rapoport effect.

The rescue effect is a phenomenon whereby local extinction of a population, often in marginal or sink habitats, is prevented because of immigration of individuals from source or high-quality habitats. Because smaller ranges, which are differentially situated in the tropics as a consequence of the Rapoport effect, have greater perimeter to area ratios, they are predisposed to having greater rescue effect areas relative to range areas. This differentially inflates species richness in tropical areas, generating the latitudinal gradient of diversity.

The generality of Rapoport's rule, as well as the degree to which empirical patterns are generated by the hybrid mechanisms embodied in rescue and Rapoport effects, is controversial. The Rapoport effect has been documented for a diversity of taxa (mammals, reptiles, amphibians, fish, crayfish, amphipods, mollusks, and trees) in aquatic and terrestrial environments and quickly has become engrained as the explanation for species diversity gradients in a variety of ecology textbooks. Additional circumstantial evidence was derived from the observation that taxa, which do not show the general latitudinal gradient in richness, do not adhere to Rapoport's rule, suggesting that both patterns had a shared mechanistic basis (Stevens, 1989). Nonetheless, a growing body of evidence suggests that the pattern is far from universal (Rohde *et al.*, 1993; Lyons and Willig, 1997). Moreover, reanalyses of data on marine mollusks that was used to corroborate the Rapoport effect (Stevens, 1989) failed to produce the same patterns in a subsequent study, even though the methods were the same in both studies (Roy *et al.*, 1994). In addition, New World bats and marsupials (Lyons and Willig, 1997; Willig and Lyons, 1998), as well as nonmigratory marine teleosts from surface waters (Rohde *et al.*, 1993), each exhibit strong latitudinal gradients in diversity but do not adhere to Rapoport's rule. Hence, occurrences of latitudinal gradients in diversity do not have one-to-one correspondence with the existence of a Rapoport effect.

Recent simulation models provide added insight into the phenomenon of Rapoport's rule. The three commonly used methods (i.e., Stevens', midpoint, and most-distal point) for assessing a Rapoport effect suffer from serious limitations. Stevens' method is problematic because of a lack of independence associated with counting the same species multiple times in the same analysis (Rohde *et al.*, 1993). Midpoint and most-distal point methods suffer from severe mathematical biases—the bounded nature of continents or oceans predisposes correlations between range size and latitude even when ranges are distributed stochastically with respect to latitude (Colwell and Hurtt, 1994; Lyons and Willig, 1997). Finally, a comprehensive set of simulation models (Taylor and Gaines, 1999) suggests that the Rapoport effect causes a latitudinal pattern in species richness, but the gradient is opposite of the pervasive pattern found in nature in that species richness increases with increasing latitude. Moreover, incorporation of a rescue effect into the model so that it reflects the Rapoport–rescue mechanism still fails to rescue the hypothesis; the predicted pattern remains a decrease in richness toward the tropics. Only the incorporation of competitive effects to either the simulation model based on the Rapoport effect alone or to the combined Rapoport–rescue mechanism produces latitudinal gradients that are consistent with real-world patterns, and in both scenarios, communities must be saturated at equilibrial compositions.

D. Geometric Constraint Hypotheses

The ubiquity of the latitudinal increase in diversity with decreasing latitude stimulated the search for a predominant mechanism effecting variation in richness. The rationale was essentially that because almost all taxa on all continents as well as in the oceans followed the same pattern, and the pattern has persisted through geological time, a single mechanism must be causing it. Interestingly, the search for a predominant mechanism during the past 25 years has generated increasingly more hypotheses rather than leading to consensus or a synthetic understanding of mechanisms producing the universal pattern.

Insights during the past 5 years concerning the nature of geographic constraints in affecting patterns of species distribution document clearly that modal patterns of diversity peaking in the tropics can be a consequence of the bounded nature of terrestrial and aquatic habitats. Indeed, both simulation (Colwell and Hurtt,

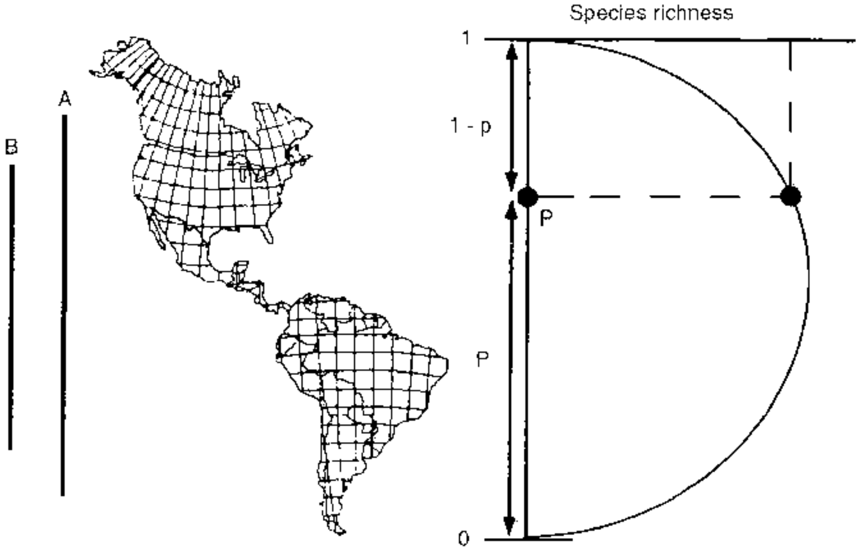


FIGURE 8 Graphical representation (parabola) of the gradient of species richness that arises from the random placement of species ranges within the latitudinal bounds of the New World, scaled from 0 in the south to 1 in the north (modified from Willig and Lyons, 1998). The number of species at any latitude is determined by its proportional distance (p) from the southern boundary and is given by $(2p - 2p^2)S$. The vertical lines labeled A and B represent the latitudinal extents of bats and marsupials, respectively.

1994; Pineda and Caswell, 1998) and analytical (Lees *et al.*, 1999; Willig and Lyons, 1998) null models suggest that species richness of a biota should increase toward the center of a shared geographic domain in a quasi-parabolic or parabolic fashion as a consequence of the random placement of species ranges [the mid-domain effect of Colwell and Lees (1999)]. Three null models have been developed that differ in the manner in which species ranges are randomized. Unlike the other mechanisms proposed to account for latitudinal gradients, which only suggest qualitative increases in richness with decreasing latitude, geometric constraint models make quantitative predictions concerning the form of the latitudinal gradient so that expected values for richness occur for each latitude and can be compared to empirical data.

In the fully neutral model (Colwell and Hurtt, 1994; Willig and Lyons, 1998), the placement of termini for each species distribution is random, with the geometric constraint that they occur within a shared boundary or domain. As a consequence, the number of species at any point within the domain (S_p) is only related to the proportional distance of that point from the boundary (p) and the number of species in the species pool (S), and is given by $S_p = 2p(1 - p)S$. The model (Fig. 8) is an incarnation of both the two-hit broken stick model of MacArthur and the binomial distribution. In essence, if the latitudinal domain of a biota is rescaled to range from 0 in the south to 1 in the north, then the likelihood of a species range overlapping any point P that is exactly p from the southern terminus (and hence $1 - p$ from the northern terminus) of the domain is

$$Pr(P) = 1 - p^2 - (1 - p)^2 = 2p - 2p^2,$$

where p^2 is the proportion of species whose northern and southern boundaries lies to the south of P, and $(1 - p)^2$ is the proportion of species whose northern and southern boundaries lie to the north of P. The functional form of the distribution of $Pr(P) = 2p - 2p^2$ is a parabola that peaks at 0.5, and as a consequence the average size range of species in the biota is half the extent of the domain (0.5).

The other two models have constraints concerning either the range size distributions of the biota or the distributions of midpoints within the domain. In the second model, the placement of ranges within the domain is constrained such that the simulated ranges have a size distribution exactly equal to that of the biota of interest. Essentially, the size of a species range limits the options for the feasible placement of midpoints: Species with broad distributions must have midpoints located near the center of the domain, whereas species with increasingly narrow ranges can have midpoints located ever more distant from the center of the domain. Hence, randomization of location for the midpoint of each species range is constrained to only occur within a subset of the domain, and within this subdomain its position is determined from a uniform random distribution. This produces a quasi-parabolic curve that becomes increasingly flattened as mean range size diverges from 0.5 (the mean value in the totally null model). The third model constrains the distribution of midpoints to be exactly the same as that in the empirical data. It essentially allows the extent of each species range to vary in a uniform random manner with the only constraint being that its distributional midpoint not change. Although it exists as a simulation null model, its analytical analog has not been developed (Colwell and Lees, 1999).

Empirical support for geometric constraints is limited but increasing. In a comparison of empirical latitudinal gradients of richness with predictions of a geometrically constrained null model in the New World, Willig and Lyons (1998) were able to account for 69–95% of the variation in species richness for bats and marsupials (Fig. 9). Nonetheless, systematic deviations from the null distribution were observed for both taxa. The null model overestimated bat species richness near the edges of the domain and underestimated richness near the center of the domain. In contrast, the null model consistently overestimated species richness for marsupials at all latitudes. Although both taxa gradually increase in species richness toward the tropics, the manner in which they deviated from the predictions of the null model is in sharp contrast, and the residuals were not related to the area or the width of the continent at each latitude.

The geometric model constrained to conform to empirical range size distributions accounted for the spatial distribution of endemic rain forest taxa (e.g., butterflies, frogs, rodents, tenrecs, chameleons, and birds) in Madagascar far better than did mechanisms related to area, elevation, temperature, precipitation, habitat diversity, or productivity (Lees *et al.*, 1999). Richness peaked at or near the midpoint of domains even though values for numerous environmental characteristics did not do so. This was true for the one-dimensional domain of latitude (the model accounts for 85% of variation in richness) as well as for the two-dimensional domain defined by latitude and longitude (the model accounts for 75% of the variation in richness).

In an analogous fashion, applications of geometric constraint models to elevational distributions of birds (Rahbek, 1997) or bathymetric distributions of gastropods and polychaetes (Pineda and Caswell, 1998) have led to considerable insight concerning diversity gradients. This is particularly relevant because elevational gradients are thought to recapitulate latitudinal patterns and mechanisms. For New World tropical birds, a peak in species richness occurred at an intermediate elevation, after controlling for the effects of area. Indeed,

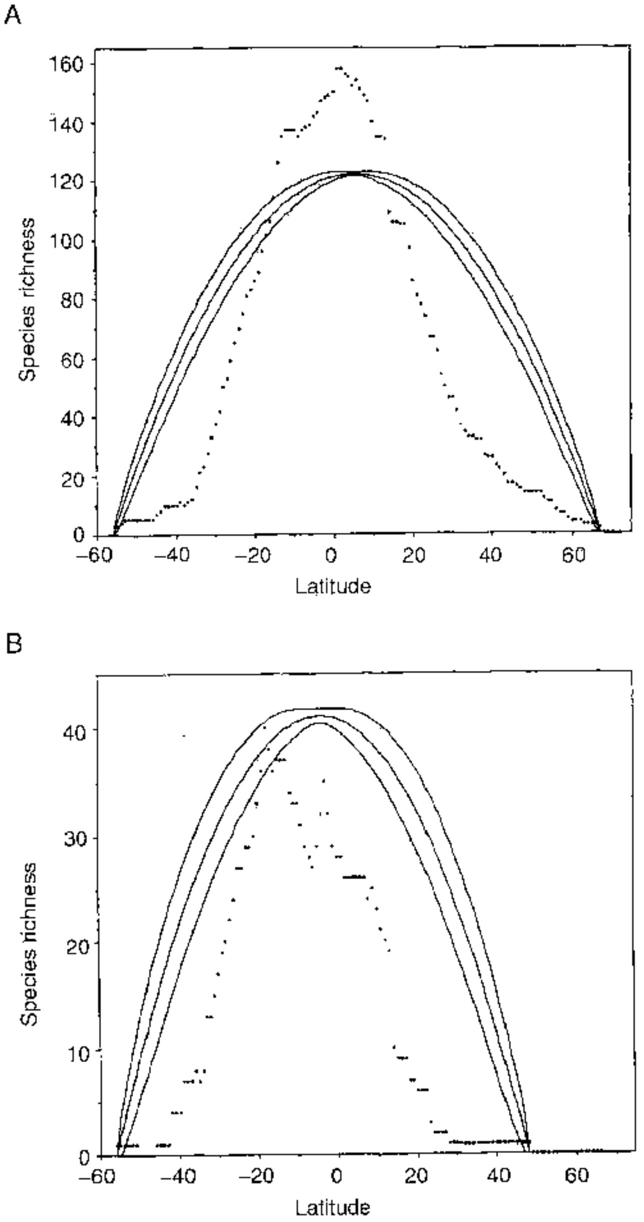


FIGURE 9 Comparison of the empirical latitudinal gradient of species richness (dots) with the predicted values (outer lines represent 95% confidence bands, and inner line represents predicted value) generated by a null model constrained only by the distributional bounds of the fauna for bats (A) and marsupials (B) in the New World (modified from Willig and Lyons, 1998).

a geometric model without constraints related to the elevational distribution of midpoints or extents, rather than other hypotheses, accounted for a significant amount of the elevational variation in species richness. Nonetheless, the actual midelevational peak in richness was depressed compared to the exact midelevation. In contrast, a similar model applied to marine invertebrates in the North Atlantic failed to account for salient features of the bathymetric gradient in richness. Although the location of a midgradient peak in richness did correspond between the model and the empirical pattern for gastropods, the model failed to account for the curvature and kurtosis of the empirical diversity gradient. For polychaetes, the model predicted the value for maximal richness. Nonetheless, the location of the peak along the depth gradient and the curvature of the relationship did not correspond between empirical data and model predictions.

Clearly, geometric constraints rather than evolutionary, environmental, or historical factors can produce latitudinal gradients that share many quantitative features with empirical patterns. Random processes may predispose various biotas to produce gradients with peaks in richness at midlatitudes. The challenge to macroecology is to understand the mechanisms that result in deviations from such null models, as well as to account for peaks in diversity and define the form of the diversity–latitude relationship per se.

IV. AMPLIFICATION AND SYNTHESIS

Based on first principles, Kaufman *et al.* (2000) provided a synthetic model to explain the latitudinal gradient in diversity based on variation in abiotic stress, productivity, and biotic interactions. The model is comprehensive because it simultaneously considers and makes predictions about other macroecological patterns, such as latitudinal gradients in range size, habitat specificity, and species dominance. Central to their model is the premise that abiotic stress increases with increasing latitude. The stress is a result of three interacting phenomena that vary with latitude as a consequence of the spherical nature of the earth and the fact that it rotates on a tilted axis with respect to the sun. Solar radiation is the ultimate source of energy for all food webs, and the daily input of energy per unit area decreases from the tropics to the poles. Similarly, the two essential metabolic pathways that dictate energy transformation in the earth's living systems—photosynthesis and respiration—are temperature dependent, and average daily temperature decreases from tropical to polar latitudes, increasing the costs of executing all life-sustaining physiological processes. Finally, intra-annual variation in both temperature and solar insolation increases from tropical to polar regions. Together, these phenomena result in increasingly energy-poor and stressful environments toward high latitudes which reduces the richness of local communities and assemblages. Increased costs of maintaining metabolic rates combined with reduced inputs of energy in extratropical regions result in population densities of a species diminishing at high latitudes, whereas increased biotic pressures resulting from interspecific interactions reduce densities at the tropical edge of a species boundary. Consequently, species are expected to have modal geographic distributions of population size. Second-order biotic feedbacks (e.g., parasitism, disease, and diffuse competition) are hypothesized to further increase the costs of survival where richness is high, thereby truncating the extent of species distributions and the abundance of local populations in tropical regions. This would generate a Rapoport effect and increase the rate of species turnover (beta diversity) and diminish dominance toward the tropics. Although data are insufficient to test all of the patterns predicted by the synthetic model, and there is controversy surrounding the ubiquity of the Rapoport effect, for heuristic reasons alone the conceptual model provides a springboard from which enhanced understanding of the causes and consequences of the latitudinal gradient may be forthcoming.

V. ASSESSMENT

The ontogeny of theory can be viewed from a variety of perspectives that deal with the detection of patterns, the linkage of patterns to particular mechanisms, and ultimately the integration of those constructs to other theories in the discipline (Pickett *et al.*, 1994). The theory of latitudinal gradients of diversity has matured considerably in the past 5 years. The general patterns of latitudinal increase are well documented from an empirical perspective. In addition, the manner in which particular mechanisms could affect patterns of diversity has become clearer. Elements of the theory have been used to understand other gradients of diversity, such as those related to elevation or depth. Finally, latitudinal patterns of diversity are being integrated with other broad-scale patterns concerning the dominance and turnover of species as well as to latitudinal patterns of range size and abundance. This represents a significant advancement in understanding and integration.

Although few of the hypotheses postulated to affect the latitudinal gradient in diversity have been eliminated in a conclusive manner, research concerning many of the mechanisms appears more likely to advance theory in the near future. Indeed, recent synthetic works have focused on them to a great extent. Understanding the contexts and degrees to which area, climatic variability or stress, geographic constraints, productivity, temperature, and their interactions mold the latitudinal gradient in diversity remains a challenge for the next decade of scientists to address in a synthetic manner.

See Also the Following Articles

BIODIVERSITY-RICH COUNTRIES • BIOGEOGRAPHY, OVERVIEW • HOTSPOTS • SPECIES-AREA RELATIONSHIPS • TROPICAL ECOSYSTEMS

Bibliography

Brown, J. H. (1995). *Macroecology.* Univ. of Chicago Press, Chicago.

Colwell, R. K., and Hurtt, G. C. (1994). Nonbiological gradients in species richness and a spurious Rapoport effect. *Am. Nat.* 144, 570–595.

Colwell, R. K., and Lees, D. C. (1999). The mid-domain effect: Geometric constraints on the geography of species richness. Submitted for publication.

Currie, D. J. (1991). Energy and large-scale patterns for animal- and plant-species richness. *Am. Nat.* 137, 27–49.

Fischer, A. G. (1960). Latitudinal variations in organic diversity. *Evolution* 14, 64–81.

Gaston, K. J. (Ed.) (1996). *Biodiversity: A Biology of Numbers and Difference.* Blackwell, London.

Kaufman, D. M. (1995). Diversity of New World mammals: Universality of the latitudinal gradients of species and bauplans. *J. Mammal.* 76, 322–334.

Kaufman, D. M. (1998). The structure of mammalian faunas in the New World: From continents to communities. Unpublished doctoral dissertation, University of New Mexico, Albuquerque.

Kaufman, D. M., Sax, D. F., and Brown, J. H. (1999). The latitudinal diversity gradient: Toward a unified theory. Unpublished manuscript.

Lees, D. C., Kremen, C., and Adriamampianina, L. (1999). A null model for species richness gradients: Bounded range overlap of butterflies and other rainforest endemics in Madagascar. *Biol. J. Linnean Soc.* 67, 529–584.

Lyons, S. K., and Willig, M. R. (1997). Latitudinal patterns of range size: Methodological concerns and empirical evaluations for New World bats and marsupials. *Oikos* 79, 568–580.

Lyons, S. K., and Willig, M. R. (1999). A hemispheric assessment of scale-dependence in latitudinal gradients of species richness. *Ecology*, 2483–2491.

Pickett, S. T. A., Kolasa, J., and Jones, C. A. (1994). *Ecological Understanding: The Nature of Theory and the Theory of Nature.* Academic Press, San Diego.

Pineda, J., and Caswell, H. (1998). Bathymetric species-diversity patterns and boundary constraints on vertical range distributions. *Deep-Sea Res. II* 45, 83–101.

Putman, R. J. (1994). *Community Ecology.* Chapman & Hall, London.

Rahbek, C. (1997). The relationship between area, elevation and regional species richness in Neotropical birds. *Am. Nat.* 149, 875–902.

Rohde, K. (1992). Latitudinal gradients in species diversity: The search for the primary cause. *Oikos* 65, 514–527.

Rohde, K. (1997). The larger area of the tropics does not explain latitudinal gradients in species diversity. *Oikos* 79, 169–172.

Rohde, K. (1998). Latitudinal gradients in species diversity. Area matters, but how much? *Oikos* 82, 184–190.

Rohde, K., Heap, M., and Heap, D. (1993). Rapoport's rule does not apply to marine teleosts and cannot explain the latitudinal gradient in species richness. *Am. Nat.* 142, 1–16.

Rosenzweig, M. L. (1995). *Species Diversity in Space and Time.* Cambridge Univ. Press, Cambridge, MA.

Rosenzweig, M. L., and Sandlin, E. A. (1997). Species diversity and latitude: Listening to area's signal. *Oikos* 80, 172–175.

Roy, K., Jablonski, D., and Valentine, J. W. (1994). Eastern Pacific molluscan provinces and the latitudinal diversity gradient: No evidence for "Rapoport's rule." *Proc. Natl. Acad. Sci. USA* 91, 8871–8874.

Stevens, G. C. (1989). The latitudinal gradient in geographical range: How so many species coexist in the tropics. *Am. Nat.* 133, 240–256.

Stevens, G. C. (1992). The elevational gradient in altitudinal range: An extension of Rapoport's latitudinal rule to altitude. *Am. Nat.* 140, 893–911.

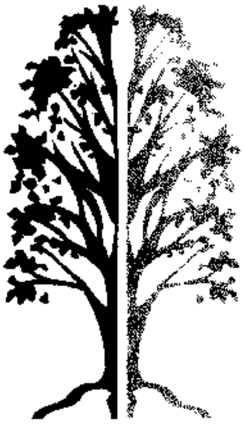
Stevens, R. D., and Willig, M. R. (1999). Size assortment in New World bat communities. *J. Mammal.* 80, 644–658.

Taylor, P. H., and Gaines, S. D. (1999). Can Rapoport's rule be rescued? Modeling causes of the latitudinal gradient in species richness. *Ecology*, 2474–2482.

Terbrogh, J. (1973). On the notion of favorableness in plant ecology. *Am. Nat.* 107, 481–501.

Voss, R. S., and Emmons, L. H. (1996). Mammalian diversity in Neotropical lowland rainforests: A preliminary assessment. *Bull. Am. Mus. Natural History* 230, 1–115.

Willig, M. R., and Lyons, S. K. (1998). An analytical model of latitudinal gradients of species richness with an empirical test for marsupials and bats in the New World. *Oikos* 81, 93–98.



LIFE HISTORY, EVOLUTION OF

Derek Roff
McGill University

GLOSSARY

genotype The genetic constitution of an organism.
heterozygosity The presence of alternate alleles at a given locus in a diploid organism (e.g., A_1A_2).
homozygosity In a diploid organism, the presence of the same alleles at a given locus (e.g., A_1A_1).
iteroparity Repeat breeding (see *semelparity*).
semelparity Breeding once and dying; sometimes called "big bang" reproduction.

THE ANALYSIS OF LIFE HISTORY evolution includes any trait that impinges on the reproductive success of an organism. More specifically, life history evolution is typically concerned with the evolution of the age and size at first reproduction, reproductive effort, clutch size, and propagule size. While many analyses focus only on a single trait, it should be remembered that selection acts on fitness (as defined later) and not solely on single traits. Therefore, the appropriate framework for the analysis of life history evolution is the whole suite of traits that interact to determine the fitness of an organism. The analysis of components of fitness is appropriate in many circumstances but the limitations of such an analysis must always be remembered.

I. INTRODUCTION

Plants and animals show profound variation in all aspects of their life histories, which include age at maturity, age-specific fecundity, survival rate, size at birth, and so on. This variation is evident at both the inter- and intraspecific levels. For example, at the interspecific level, species of flatfish range in size from 2 cm-long tropical species that reproduce within their first year of life to behemoths such as the Pacific Halibut (*Hippoglossus hippoglossus*), which exceed 200 cm and take over 10 years to mature. Though the range in variation within a species is not as dramatic as between species, it is still impressive, as illustrated by variation in the flatfish, *Hippoglossoides hippoglossoides*. In this species maturation occurs at age 3 years at a length of 20 cm in populations off the coast of Scotland while the same species requires 15 years to reach maturity at a length of 40 cm on the Grand Banks of Newfoundland. Longevity and maximum size are equally different in the two populations, Scottish fish reaching a maximum length of 25 cm and an age of 6 years, compared to 60 cm and 20+ years on the Grand Banks. Similar observations on variation in life history characteristics could be made in most taxa. But though the diversity of life histories is readily apparent, attempts to understand its origin and maintenance are still in their infancy.

Because fitness varies quantitatively among different life histories, a necessary tool for analysis is mathematical modeling. An influential factor encouraging the use of mathematical investigation into life history variation was Lamont Cole's 1954 paper, "The Population Consequences of Life History Phenomena," which set out the basic mathematical framework by which the consequences of variation in life history traits can by analyzed. Cole's paper ushered in an era of research predicated on the integration of mathematics and biology in the study of the evolution of life history patterns.

In his review, Cole analyzed how changes in demographic attributes, such as the age at first reproduction, influenced the rate of increase of a population. Cole's paper gained widespread notice because of an apparent paradox with respect to the value of semelparity versus iteroparity: "*For an annual species, the absolute gain in intrinsic population growth which could be achieved by changing to the perennial reproductive habit would be exactly equivalent to adding one individual to the average litter size*" (Cole, 1954, p118, Cole's italics). Thus, an annual species with a clutch size of 101 would increase in numbers as fast as a perennial that produces 100 young every year forever. There is obviously something amiss with this result, for perennials are common. While there is good evidence that survival and reproduction are negatively correlated it seems highly unlikely that perennial species are committing so much energy into reproduction that they cannot produce one more offspring. Cole's paradox derives from a failure to consider the consequences of juvenile and adult survival rates; specifically Cole implicitly assumed that the survival rate of juveniles is the same as that of the adults. In general, juvenile survival is lower (often by a great amount) than that of the adults and this means that the reproductive output of the annual must be greater than one for the annual to have the same gain in rate of increase as a perennial, the precise amount depending on the mathematical details of the model. Cole's paradox illustrates the need to formulate an analysis of life history variation within a mathematical framework where the biological assumptions are explicitly stated. Since Lamont Cole, the theoretical and experimental analysis of life history evolution has made enormous strides using two different analytical perspectives.

II. TWO FRAMEWORKS FOR ANALYSIS

The evolution of life history variation has been approached via two different modes of analysis, here termed phenotypic models and genetic models.

A. Phenotypic Models

In this approach no attempt is made to model the genetic basis of traits: it is simply assumed that there exists sufficient genetic variation that evolution is not constrained by genetic architecture. Is this a reasonable assumption? Most life history traits—such as age at first reproduction, fecundity, and survival—are not determined by simple Mendelian mechanisms such as single locus, two allele systems. More generally they show continuous variation, or if dichotomous can be best interpreted under a threshold model in which there is an underlying continuously varying trait. Nevertheless, polygenic determination is not essentially different from the simpler case of a single locus with two alleles. In the one locus case the phenotype is determined by the additive action of the alleles, or one allele may show dominance. When one allele is dominant, the effect on the phenotype can be statistically decomposed into additive and dominance components. With two or more loci controlling a trait there may be interaction between loci. These interactions, termed epistasis, can produce a wide range of responses but are generally assumed to be absent (because most variation can be subsumed under the additive or dominance portion of genetic variance, this assumption is not as unreasonable as it might appear).

The average similarity between parent and offspring can be measured by the linear regression of mean offspring value (Y) on midparent value (X),

$$Y = (1 - h^2)m + h^2X \tag{1}$$

where m is the population mean of the parents and the slope of the regression, h^2, is termed heritability in the narrow sense. More generally, heritability is defined as follows:

$$h^2 = \frac{V_A}{V_P} = \frac{\text{Additive genetic variance}}{\text{Total phenotypic variance}} \tag{2}$$

Heritability in the narrow sense, or generally simply heritability, must not be confused with heritability in the broad sense, V_G/V_P, which is a measure of the overall variance in the trait attributable to all genetic influences.

TABLE I

A Summary of Heritability Estimates for Nondomestic Animals[a]

Comparison	Life history	Behavior	Physiology	Morphology
Drosophila	0.12	0.18	ns	0.32
All animals	0.26	0.30	0.33	0.46
Medians	0.26	0.37	0.31	0.51

[a] Data are from Roff and Mousseau (1987) and Mousseau and Roff (1987).

Obviously h^2 varies between 0 and 1: at 0 there is no resemblance between parent and offspring due to the additive effects of genes, while at 1 the mean offspring value equals the midparent value. The importance of heritability for evolutionary theory is obvious: the higher the value of h^2, the faster genetic changes in the population can occur. Most important from the point of view of life history theory, if heritability is zero for a particular trait or traits under consideration, then the optimal combination cannot be attained because the effects of selection on parents are not manifested in the offspring. Heritability estimates for different types of traits are shown in Table I. While heritabilities for life history traits are typically lower than those less directly related (in general) to fitness, there is nevertheless sufficient genetic variation to permit rapid evolutionary change even under modest selection pressure.

While the assumption of sufficient additive genetic variation is reasonable with respect to the equilibrium conditions, the genetic architecture can considerably influence evolutionary trajectories and hence the phenotypic approach is contingent on the state having attained an equilibrium. The method seeks to construct the fitness surface and hence the optimal combination of trait values.

1. Optimality Modeling

The concept of tradeoffs is central to present theories of the evolution of life history traits, for tradeoffs limit the scope of variation. Within the set of possible combinations there will be at least one combination that exceeds all others in fitness. Optimality analysis assumes that natural selection will drive the organism to that particular set. To initiate an analysis using the principle of optimality, we must designate some parameter to be optimized. In the present case we assume that there is some measure of fitness that is maximized by natural selection. The second step is to construct a set of rules that define the life history pattern of the organism, hypothetical or real, under study. Within these rules there will exist a variety of possible life histories; the optimal life history is that which maximizes fitness.

What do we do if the predicted life history does not correspond to that which is observed? The first point to note is that the principle of optimality is not under test. Failure to get a correct prediction is not taken as evidence that fitness is not being maximized, but it is taken to imply that the model is deficient. Having found that the initial model does not work we inquire into the assumptions of the model, namely the rules that define the range and scope of life history variation. These rules are changed either arbitrarily or based on further observation until agreement is gained between prediction and observation. Having found congruence, we are not able to say that therefore the component relationships are correct. The underlying components must be independently verified; a model is only the logical outcome of a set of interactions and the onus is on the person specifying those rules to demonstrate that they are indeed valid. Note again, that the assumption that fitness is being maximized is not under test, except to the extent that the particular measure chosen may be inappropriate. This does not mean that we assume that all traits and trait combinations are the result of adaptive evolution. But we do choose those that we have a priori reason to suppose are under selection.

A primary purpose of optimality modeling is to organize a program of experimentation and data collection. An adequate fit of a model to data gives us reassurance that a sufficient number of factors have been taken into account. Nevertheless, the validity of a model is continually under question and is challenged by the addition of more information. The more tests the model survives the greater the assurance that it is realistically capturing the important elements that determine the set of traits being studied.

2. Game Theory

Game theory is really a subset of optimality modeling; it is appropriate when interactions are frequency dependent. The approach comprises two essential elements. First, it is assumed that particular patterns of behavior will persist in a population provided no mutant adopting an alternate behavior can invade. Such stable combinations are termed evolutionarily stable strategies (ESS). The concept of the ESS is not unique to game theory: the maximization of fitness measures in optimality models are all ESSs within the context in which they are appropriate. Second, for each type there must be an assigned gain or loss in fitness when this type interacts with another individual. From this payoff matrix we compute the expected payoff for each behavior.

For two behaviors to be evolutionarily stable, their fitnesses must be equal.

B. Genetic Models

Genetic modeling proceeds by first defining the genetic mechanism determining the phenotypic trait and then cranks the model through the appropriate mathematical machinery to obtain the equilibrium trait values. The critical problems are the correct definition of the genetic architecture and the estimation of the selection gradients.

1. Simple Mendelian Models

Consider a trait whose expression is governed by a single locus with two alleles, A_1, A_2. A selection coefficient (= fitness) representing their relative contribution to the next generation can be assigned for each of the three genotypes (Table II). The change in the frequency of A_1, p, is then easily obtained as

$$\Delta p = \frac{pq[p(w_{11} - w_{12}) + q(w_{12} - w_{22})]}{\bar{w}} \qquad (3)$$

At equilibrium $\Delta p = 0$, and hence it is a trivial matter to predict the frequencies of the three genotypes at equilibrium and hence the mean trait value. Complexity can be added by making the fitnesses a function of density or frequency, but this does not change the basic mathematical approach.

Most traits of ecological interest, fecundity, age at maturity, clutch size, egg size, and so on are continuous in character. Even traits that appear dichotomous, such as liability to disease, wing dimorphism, diapause, and sex ratio, are best understood as being the result of some underlying continuously varying factor exceeding or not attaining a threshold for the expression of the trait. The expression of traits that show continuous variation are not, in general, the result of a single gene, nor two genes, but a large number of genes that acting additively produce a continuous spectrum of phenotypes. The analysis of such traits is the domain of quantitative genetics. This is largely a statistical approach to genetic variation and is founded on a mathematical analysis of variation rather than an understanding of how groups of genes interact to determine a particular trait.

TABLE II

The Derivation of the Optimal Trait Value in a Simple Single-Locus Model

	A_1A_1	A_1A_2	A_2A_2	Sum
entry align="left" valign="top" colname="c1"Ratios before selection	p^2	$2pq$	q^2	1
Fitness	w_{11}	w_{12}	w_{22}	
Ratios after selection	$p^2 w_{11}$	$2pqw_{12}$	q^2w_{22}	w

2. Quantitative Genetic Models

The concept of heritability has already been introduced. For nonzero heritabilities, the rate at which the appropriate combination of trait values can be realized depends in part on the value of the heritability and the intensity of selection, measured as the selection differential, S, the difference between the mean of the population and the mean of the selected parents. For a single trait it can been shown from Equation 1 that the response to selection (the difference between the mean population values in parental and offspring generations), R, is equal to h^2S. Since selection changes the gene frequencies, heritability must change at each generation of selection. This problem is avoided in the application of quantitative genetic theory to natural populations by assuming that selection is weak and population size large, thereby permitting mutation to replace variation eroded by selection. Whether this assumption is reasonable can only by answered empirically and at present there are insufficient data to draw any meaningful conclusion.

Traits are typically not inherited as separate independent units; rather, phenotypic values of several traits are typically controlled in part by a common set of genes. Thus selection on one trait will produce a change not only in the selected trait but also traits that are genetically correlated by virtue of shared genes (genetic correlations can also arise through linkage disequilibrium, but these are transitory and are ignored here). Selection on two traits, X and Y, will produce standardized (in phenotypic standard deviation units) responses R_X and R_Y,

$$\begin{aligned} R_X &= \beta_X h_X^2 + \beta_Y h_X h_Y r_A \\ R_Y &= \beta_Y h_Y^2 + \beta_X h_X h_Y r_A \end{aligned} \qquad (4)$$

where r_A is the genetic correlation between the two traits and β_X, β_Y are the standardized selection differentials on X and Y, respectively. The preceding equation can be more conveniently written in matrix notation, for which the typical notation is $\Delta\bar{z} = G\beta$, where $\bar{z}$ is a vector of

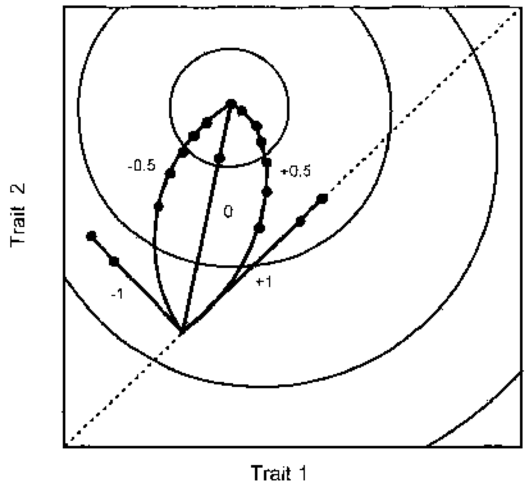


FIGURE 1 Hypothetical evolutionary trajectories for two traits. The solid lines define contours of equal fitness. Each dot represents the trait combination achieved after some specified number of generations. The values beside each trajectory indicate the genetic correlation between the two traits.

trait means, G is the genetic variance-covariance matrix, and β is a vector of (unstandardized) selection gradients for each character.

The evolutionary trajectories for two traits can be easily visualized by constructing a plot of isoclines of equal fitness (defined by the selection gradients) and plotting the changes in trait means each generation as illustrated in Fig. 1. If there is no genetic correlation between the two traits, the population will move rapidly to its optimal combination. For genetic correlations lying between −1 and +1, the trajectory is "warped" and if the genetic correlation is exactly ±1, the optimal combination may be unattainable (Fig. 1). The reason for this can be understood by considering the regression of the additive genetic values of trait Y on X. According to the quantitative genetic model, we have $Y = c + bX +$ error, where c and b are constants and the error term is normally distributed with mean zero. Provided the variance in the error term is not zero, any combination of Y and X is possible; hence, evolution can always move the traits to such a combination. However, if the variance of the error term is zero, which occurs when the (genetic) correlation between Y and X is ±1, then the traits are constrained to lie on the regression line. The number of cases in which r_A is exactly equal to unity is very small (see Section IV). With more than two traits evolution can be constrained if one or several eigenvectors of the genetic variance-covariance matrix are zero (singular matrix).

It is possible for there to be no response to selection in spite of both genetic variance for the trait and a phenotypic correlation with another trait under selection. The second observation is particularly important since it implies that a trait may be under selection but show no response. This can be most easily seen by relating the selection differential as a function of the selection gradient and phenotypic variance (V_{PX}) and covariance (Cov_P)

$$S_X = \beta_X V_{PX} + \beta_Y Cov_P \tag{5}$$

with a similar term for S_Y. Now suppose that (a) there is no additive genetic variance for trait Y, (b) trait Y is phenotypically correlated with X but genetically uncorrelated ($Cov_A = 0$), and (c) there is no direct selection on trait X (i.e., $\beta_X = 0$) but there is selection on trait Y (i.e., $\beta_Y > 0$). From Equation 5 it can be seen that a positive selection differential is generated ($S_X > 0$), but from equation 4 ($R_X = V_{AX}\beta_X + Cov_A\beta_Y$) it is apparent that there will be no response to this selection, even if there is additive genetic variance for X. A likely case in which this situation can arise is when trait X is a heritable trait such as fecundity and Y is a trait such as nutritional status that might have zero heritability: thus well-nourished individuals have high fecundity, which gives them an apparent selective advantage, but this is not realized because selection is acting only on the environmentally determined component of fecundity. The mechanism described earlier, or ones conceptually similar, have been proposed for the observation of directional selection without evolutionary response in breeding date in birds, clutch size in birds, and tarsus length in birds.

The theory described here assumes that the genetic covariance (i.e., the genetic correlation) remains constant. As with heritability, selection changes allelic frequencies and hence must in principle change the genetic correlations, or more generally the genetic variance-covariance matrix. The G matrix will remain constant provided selection is weak and population size large so that mutation can replace variance lost due to selection and drift.

III. TRADEOFFS: A CENTRAL FEATURE OF LIFE HISTORY ANALYSIS

Tradeoffs are essential elements of all the above approaches. In the "phenotypic approach" it is implicitly assumed that the tradeoffs are genetically determined, otherwise there could be no evolutionary response. In quantitative genetic terms, this is interpreted as a negative genetic correlation between two traits. Thus it has been supposed that to demonstrate an evolutionarily important tradeoff it is sufficient to demonstrate that the tradeoff is expressed not only as a phenotypic correlation but also as a negative genetic correlation. However, as discussed earlier, a negative genetic correlation is not, in principle, a barrier to movement anywhere in parameter space. Thus while it is necessary for a tradeoff to be expressed as a genetic correlation for it to be evolutionarily significant, this is not a sufficient demonstration that the optimal combination will be governed by the tradeoff.

Empirical investigations of tradeoffs can be placed into three categories: (a) phenotypic relationships based on field or laboratory observations of unmanipulated situations, (b) experiments in which organisms were manipulated to vary the value of one trait (e.g., manipulation of clutch size to investigate subsequent survival of adults or young), and (c) demonstration of a negative genetic correlation between two traits, obtained by sib analysis or selection. Only the last measure provides definitive proof that the tradeoff is evolutionarily significant but there is, however, merit in the second approach because it can establish the mechanism generating the tradeoff. The first approach is suspect because of the covariation of traits that could produce false conclusions if such covariation is not taken into account. For example, reproduction may be determined by condition, those in poor condition not breeding; consequently it would not be surprising to find the survival rate of nonbreeders to be less than breeders.

There is no reason to suppose that all tradeoffs between life history traits will be negative, but we should expect a significant proportion to be so. In fact, the distribution of genetic correlations between life history traits is very broad (Fig. 2), with 39% being negative but few estimated to be −1. In contrast, only 23% of genetic correlations between morphological traits are negative. Thus there do appear to be tradeoffs that will at least impede the evolution of life history traits.

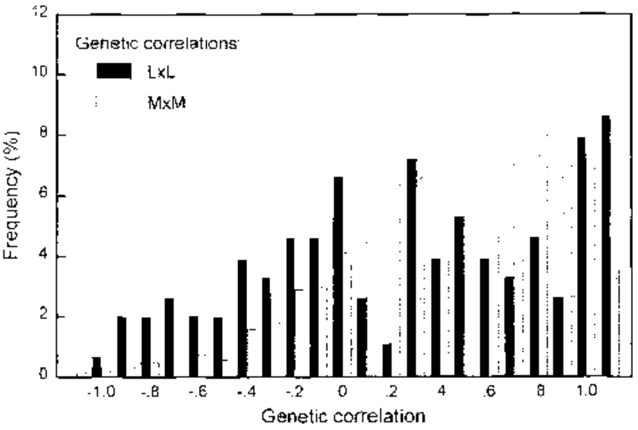


FIGURE 2 Distributions of the genetic correlation between Morphological traits (M × M) and Life History traits L × L.

IV. THE EVOLUTION OF LIFE HISTORIES

The starting point of many analyses of evolutionary change, most particularly that of optimality modeling, is the assumption that there exists some variable maximized by selection. The issue of what is being maximized has been the subject of much discussion, partly because the appropriate measure of fitness changes with circumstance and method of analysis. Broadly, measures of fitness can be separated into two groups: global measures and local measures. A global measure of fitness is one that involves the interaction of all life history components, the best example being Fisher's Malthusian parameter r. Local measures assume that maximization of a fitness component will also maximize the overall fitness of the organism: for example, a common local measure used in foraging theory is the net rate of energy intake. This is an appropriate measure if it can be shown that maximizing this rate does not detrimentally affect other components of fitness, in which case maximizing net rate of energy intake will also increase global fitness.

Local measures are generally tailored for the particular analysis under consideration, but there now exists a general consensus, and more important sound theoretical rationale, of what global measures are likely maximized by natural selection.

A. Static Environments

A population growing in an unlimited, homogeneous, and constant environment follows the simple exponential growth function

$$\frac{dN(t)}{dt} = rN(t) \qquad (6)$$

$$N(t) = N(0)e^{rt}$$

where $N(t)$ is population size at time t and r is the intrinsic rate of increase, comprising the difference between instantaneous rates of birth and death. Equation

6 can also be written as

$$N(t) = N(0)(e^r)^t = N(0)\lambda^t \qquad (7)$$

The symbol λ is called the finite rate of increase and is sometimes used instead of r.

Suppose there are two clones with population sizes, $N_1(t)$ and $N_2(t)$, respectively, the first with an intrinsic rate of increase of r_1 and the second with r_2, with $r_1 > r_2$. The ratio of population sizes after some time t, given that both clones start with the same population size is,

$$\frac{N_1(t)}{N_2(t)} = e^{(r_1 - r_2)t} \qquad (8)$$

It is clear that as time progresses this ratio will increase, clone 1 becoming numerically more and more dominant in the combined population. This conclusion does not depend on the two clones beginning with the same population size: differences in starting condition simply accelerate or retard the rate at which clone 1 increases in frequency relative to clone 2.

For these two clones, an appropriate measure of fitness is r, since the frequency of the clone with the highest value of r will increase toward unity. Thus any mutation in a set of clones that increases r by changing rates of birth or death will increase in frequency in the population. There are no difficulties in assigning r as a measure of fitness in the preceding circumstances. Difficulties arise, however, when sexual reproduction and age structure are introduced. Suppose we have a random mating population in which a mutation arises that increases birthrate or decreases death rate. Since the mutant will initially be rare in the population, its fate can be ascertained by considering the birthrates and death rates of the heterozygote alone. If the heterozygote's rate of increase is enhanced, the mutation will increase in frequency in the population, but its ultimate fate depends on the relative birthrates and death rates of the homozygotes and heterozygotes bearing the mutant allele. If the homozygote carrying both mutant alleles has a higher birthrate or a lower death rate than the heterozygote, the mutant allele will eventually be fixed in the population; otherwise the population will reach a stable polymorphism.

The general assumption, stemming from the work of Fisher, has been that r can be associated with genotypes that follow particular life histories and that selection will favor that genotype with the highest value of r. In an age-structured population, the rate of increase is obtained by solving the characteristic equation

$$\int_0^\infty e^{-rx} l(x)m(x)dx = 1 \qquad (9)$$

where $l(x)$ is the probability of surviving to age x and $m(x)$ is the number of female births at age x. The discrete time equivalent of this is

$$\sum_{x=1}^{\infty} e^{-rx} l(x)m(x) = 1 \qquad (10)$$

Note that in the discrete version the initial age is subscripted as 1 not 0. The important issue to be considered is the fate of a mutant that increases r. Charlesworth demonstrated that to a rough approximation the rate of progress of a rare gene eventually becomes directly proportional to its heterozygous effect on r. In this case, the probability of survival of a mutant gene of small effect in a near-stationary population is largely determined by its effect on r. Following a more detailed analysis, Charlesworth (1994) concluded

> that for the case of weak selection and random mating with respect to age, the intrinsic rate of increase of a genotype or, more generally, the mean of the male and female intrinsic rates, provides an adequate measure of fitness in a density-independent and constant environment.

Lande tackled the problem of applying quantitative genetic theory to the evolution of r in a population. Assuming weak selection, large population size, and a constant genetic variance-covariance matrix, Lande showed that in a constant selection regime, life history evolution continually increases the intrinsic rate of increase of a population until an equilibrium is reached. The intuitive appeal of r as a suitable measure of fitness thus receives qualified support from both population genetic and quantitative genetic theory.

If the population is stationary, r is zero and the characteristic equation reduces to

$$R_0 = \int_0^\infty l(x)m(x)dx = 1 \qquad (11)$$

R_0 is termed the net reproductive rate and is the expected number of female offspring produced by a female over her life span. The use of R_0 makes analysis easier and can be justified if r is very close to zero or if the density-dependent or other factors that maintain the population at some relatively stable value do not impinge on the traits under consideration. For example, population size might be controlled by density-dependent mortality in the larval stage, while the object of study is the evolution of female age at maturity. In this case we can examine the relationship between the age at maturity and fitness under the working assumption that genotypes do not differ in the characteristics of their larvae. Since population size is stable, the expected lifetime fecundity, R_0, is then the appropriate measure of fitness.

In some analyses, fitness is determined from the reproductive value at age x, $V(x)$

$$V(x) = \frac{e^{rx}}{l(x)} \int_x^{\infty} e^{-ry} l(y) m(y) dy \qquad (12)$$

The reproductive value of an individual of age x is a measure of the extent to which it contributes to the ancestry of future generations. Williams (1966) postulated that natural selection maximizes r by maximizing reproductive value at every age. A mathematically correct statement of the principle is "reproductive value at each age is maximized relative to reproductive effort at that age, although not necessarily with respect to effort at other ages."

The preceding approach has been very successful in accounting for variation in life history traits. An illustration of this is the analysis of the optimal size at first reproduction in *Drosophila melanogaster*. In *D. melanogaster* body size increases with development time and fecundity increases allometrically with adult body size, which is fixed on eclosion from the pupa. The age schedule of female births is triangular in shape (Fig. 3) and is described by the equation

$$m(x) = \frac{1}{2} c_1 L^{c_2} (1 - e^{-c_3(x-c_4)}) e^{-c_5 x} \qquad (13)$$

where the c_is are constants, L is thorax length, and x is age. Thorax length, L, scales the age schedule of reproduction, larger females producing more eggs, but does not change its position. The constant c_1 is the product of two constants: the coefficient of proportionality within the allometric relationship between length and fecundity, and the proportion of eggs that fail to hatch.

Development time, $d(L)$, scales to body size according to the relationship (Fig. 3)

$$d(L) = c_6 L^{c_7} + c_8 \qquad (14)$$

Thus development time scales allometrically with size except for the constant c_8, which represents time required for the eggs to hatch and the development within the pupa, both of these components being independent of size.

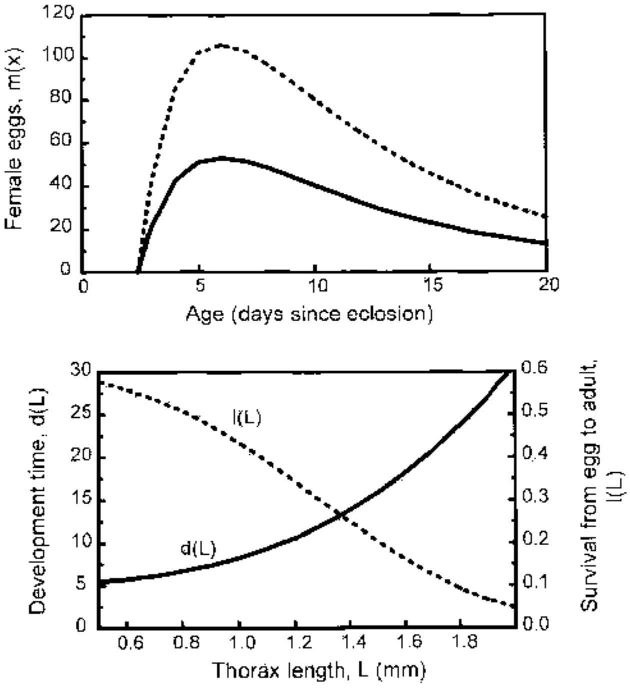


FIGURE 3 Relationship between body size and life history traits in *Drosophila melanogaster*. Upper panel: Age schedule of reproduction for small (solid line) and large (dashed line) females. Lower panel: Development time and survival from egg to adult in relation to adult thorax length.

Because information on mortality rates are so poorly known, it was assumed that instantaneous rates remain constant in the adult and larval phases at M_a and M_l, respectively. Probability of surviving the larval period, $l(L)$, is thus (Fig. 3)

$$l(L) = e^{-M_l d(L)} \qquad (15)$$

Drosophila melanogaster is a colonizing species and hence the appropriate measure of fitness is r. Combining the previous relationships, we obtain the characteristic equation

$$\sum_{x=1}^{\infty} \frac{1}{2} e^{-r[x+d(L)+c_4]-M_l d(L)-M_a(x+c_4)-c_5 c_4} c_1 L^{c_2} \qquad (16)$$
$$(1 - e^{-c_3 x}) e^{-c_5 x} = 1$$

where, for convenience age has been rescaled to begin at the first day of egg laying. Though this equation is tediously long, its solution presents no great difficulty. Briefly, the method is, first, to evaluate the series making

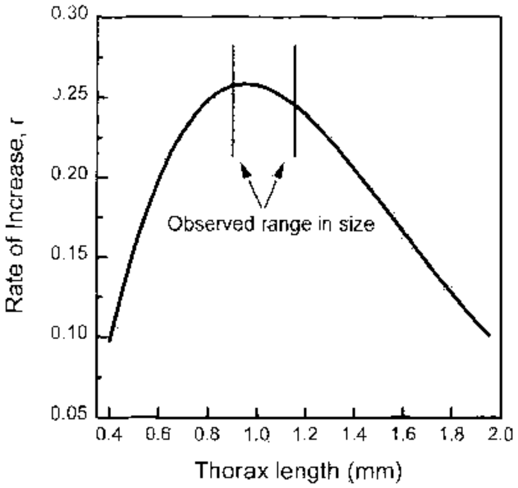


FIGURE 4 Estimated relationship between fitness, as measured by the rate of increase, and thorax length in *Drosophila melanogaster*.

use of the fact that it is a geometric progression and, second, to differentiate implicitly to obtain the optimal length at maturity. The optimum thorax length depends on adult mortality, larval mortality, and the constant initial "mortality" component (p or c_1). Figure 4 shows the relationship between r and thorax length using values obtained from laboratory stocks and estimates from wild populations. The predicted maximum, 0.95 mm, falls very nicely within the observed range in thorax length of 0.90 to 1.15 mm.

B. Stochastic Environments

Environments are typically variable in both time and space. While the assumption of a constant environment may be a reasonable first approximation in many cases, there will be many others in which variation cannot be ignored. The simplest case to consider is where the environment is stochastic and there is no cue as to future conditions.

In a stochastically, temporally fluctuating environment, the correct measure of fitness is the geometric mean of the finite rate of increase. The rationale for this measure is as follows: the size of a population after t time intervals is given by

$$N(t) = N(0)\lambda_1\lambda_2 \ldots \lambda_t = N(0)\prod_{i=1}^{i=t} \lambda_i \qquad (17)$$

The arithmetic and geometric means are

$$\text{Arithmetic mean} = \frac{1}{t}\sum_{i=1}^{i=t} \lambda_i$$
$$\text{Geometric mean} = \sqrt[t]{\prod_{i=1}^{i=t} \lambda_i} \qquad (18)$$

It is obvious that the geometric mean describes the long-term average per generation change in population size and hence is the parameter that reflects the fitness of different strategies. If r is used in place of the finite rate of increase, the appropriate measure of fitness is the arithmetic mean of r. This can be demonstrated as follows: dividing time intervals into small units we can write

$$N(t) = N(0)e^{r_1}e^{r_2} \ldots e^{r_t} = N(0)e^{\sum_{i=1}^{t} r_i} \qquad (19)$$

from which it is readily apparent that the correct measure of fitness is the arithmetic average of r.

Consider two genotypes living in an environment that comprises two types of year, "good" and "bad," each occurring with equal frequency. In "good" years, genotype A has a finite rate of increase of 2 and in a "bad" year a finite rate of increase of 0.5, while genotype B has finite rates of increase of 1 and 1.1, respectively. The arithmetic averages of A and B are 1.25 and 1.05, respectively, but the geometric averages are 1 and 1.1. Thus genotype B has the highest long-term fitness although it has a smaller arithmetic finite rate of increase. Genotype A increases more than genotype B in "good" years but suffers a greater reduction in "bad" years. The relatively high fitness of genotype B resides in the fact that although it has a smaller arithmetic average, it also has a smaller variance in its finite rate of increase. In a stochastic environment the highest fitness may be gained not by the production of a single phenotype but by a genotype producing a range of phenotypes. Such bet-hedging strategies are common, an example being the production of both diapausing and direct-developing offspring in both invertebrates and plants.

C. Predictable Environments

Most environments show predictable changes either as a result of biotic factors (e.g., overgrazing, community succession) or abiotic factors (e.g., winter, monsoons). Suppose factors that affect one or more life history characters are fully predictable—that is, given some

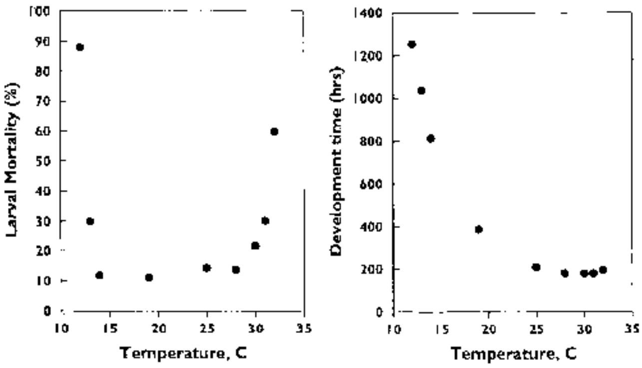


FIGURE 5 Reaction norms between two life history traits (larval mortality, development time) and temperature in *Drosophila melanogaster*.

cue E the life history characters will be changed from their present values to values that can be related to E by some function. For each case there is some optimal combination of trait values; given the predictable nature of the cue means that the organism's life history can be altered to accommodate the new conditions. Suppose, for example, that when the environmental conditions are E_i the optimal clutch size is C_i. Upon receipt of the environmental cue, the organism produces the optimal clutch size, C_i. This relationship between trait value and environment is called a norm of reaction, whether or not it is optimal. Reaction norms are ubiquitous. Two examples of reaction norms are shown in Fig. 5. A reaction norm may be adaptive or a nonadaptive (even maladaptive) physiological response to the environment. To demonstrate that a reaction norm is adaptive requires that we use the set of tradeoff functions related to the environment to predict the optimal reaction norm. Consider, for example, the question asked by Stearns and Koella (1986, p. 894): "How should an organism encountering an unavoidable stress that results in slower growth alter its age at maturity to keep fitness as high as possible despite the constraints imposed by slower growth?"

To address this question, Stearns and Koella used the following hypothetical model

1. Fecundity increases with body size, which increases with age:

$$W(x) = c_1(1 - c_2e^{-kx})$$
$$m(x) = c_3W(x) + c_4 \qquad (20)$$

where $W(x)$ is size at age x, k is the parameter determining the rate at which the asymptotic size is approached, and c_1–c_4 are constants.

2. Adult mortality rate, M_a, decreases with growth rate according to the function:

$$M_a = \frac{c_5}{k^{c_6}} \qquad (21)$$

3. Mortality rate of juveniles, M_j, is a function of the component of adult mortality, c_7, plus an amount that decreases with age at maturity, α, and growth rate, k:

$$M_j = c_5 + \frac{c_7}{\alpha^{c_8}k^{c_9}} \qquad (22)$$

Stearns and Koella assumed that fitness is maximized by maximizing r. The predicted norms of reaction between age and size at maturity for four possible scenarios are illustrated in Fig. 6:

1. Mortality rate and growth rate independent ($c_6 = 0$, $c_9 = 0$). Decreases in growth rate, k, favors an

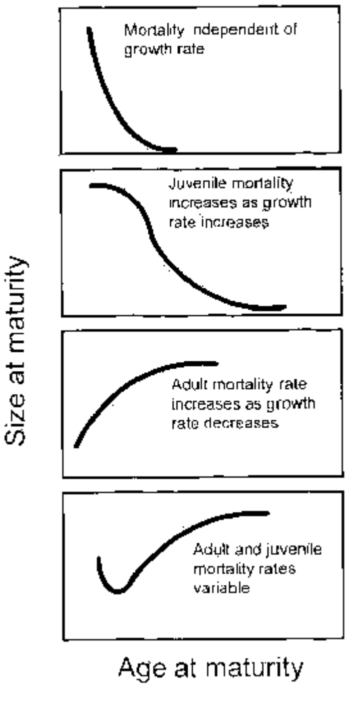


FIGURE 6 Predicted norms of reaction between size and age at maturity for the life history model described in the text.

increase in the age at maturity and a decrease in the size at maturity.

2. Juvenile mortality rate increases as growth rate increases ($c_6 = 0$. $c_0 > 0$). As with the previous case, decreases in k favor an increase in the age at maturity and a decrease in the size at maturity. The actual norm of reaction may have the same shape as case 1 or be sigmoidal.
3. Adult mortality rate increases as growth rate decreases ($c_6 > 0$, $c_0 = 0$). The optimum norm of reaction has the same shape as case 1 but is reversed, size at maturity and age at maturity being positively related.
4. The adult and juvenile rates are variable ($c_6 > 0$, $c_0 > 0$). The relationship between age and size at maturity is "keel-shaped." The portion of the curve in which size at maturity is an increasing function of age at maturity only occurs when adult mortality rates exceed juvenile mortality rates by a factor of 100 or more. Thus, generally the trajectory is determined by mortality rates of juveniles.

The message from the preceding analysis is that reaction norms can be quite complex and dependent not only on the functional relationships determining the tradeoffs but also the parameter values. The situation is made even more complex if the environment is spatially variable: in this case the appropriate fitness measure is the overall r, not the r that is characteristic of a particular environment. This requires modifying the characteristic equation to

$$\int p(h) \int l(x, h)m(x, h)e^{-rx}dxdh = 1 \qquad (23)$$

where $p(h)$ is the probability of habitat h occurring. The optimal combination of trait values can be markedly different from the average set determined by considering each habitat separately.

For reaction norms to evolve to produce the optimal response, there must be genetic variation in how individuals respond to the environment (i.e., there must be genetic variation for phenotypic plasticity). The evolution of phenotypic plasticity can be addressed using two apparently different mathematical perspectives: the character state approach and the reaction norm approach. Both approaches are actually interchangeable and each has advantages and disadvantages.

The character state approach is based on the assumption that the same trait measured in two environments can be considered as two traits that are genetically correlated. This is illustrated in Fig. 7, where each line represents a separate genotype. If the lines joining the phenotypic values in the two environments (E_1 and E_2) intersect at a common point between E_1 and E_2, the genetic correlation between the two traits is -1 (Fig. 7, panel D). If the lines intersect outside the range E_1 to E_2, the genetic correlation is $+1$ (Fig. 7, panel A; note that parallel lines also produce a genetic correlation of $+1$, since mathematically their point of intersection is at the point of infinity point beyond E_1 or E_2). The genetic correlation will differ from ± 1 if there is no common point of intersection (Fig. 8, panels B and C).

While the character state approach sees the phenotype as two points in state space, the norm of reaction approach sees a line (Fig. 8). Thus, for two environments phenotypic variation can be described as $X(E) = c_0 + c_1E + e$, where the two parameters c_0 and c_1 are viewed as traits, and e is an error term that is normally distributed with mean 0. Evolutionary change in X depends on the heritabilities of c_0 and c_1 and the genetic correlation between them. If the genetic correlation is ± 1 then the line that is genetically fixed and corresponds, as it should, to the cases A and D in Fig. 7. This description is merely an alternate formulation of the character state approach. The conceptual advantage of this approach is that it extends quite naturally to continuously distributed environmental variables such as temperature. The preceding model can clearly apply to any linear reaction norm; if the relationship between trait and environment is more complex, a more complex function can readily be substituted, for example, the quadratic $X(E) = c_0 + c_1E + c_2E^2 + e$.

The evolution to the optimal reaction norm could be constrained. Suppose, for example, the reaction norm were linear but the optimal reaction norm were quadratic: in this case the optimal reaction norm cannot be attained. If some arbitrarily complex polynomial function is assumed, then many shapes are possible and no constraint will ensue. While there are abundant data indicating that genetic variation for phenotypic plasticity of life history traits is common, there are few data suggesting that the evolution of reaction norms are constrained.

The study by Weis and Gorman (1990) on the evolution of gall size in the insect *Eurosta solidaginis* is one of the few studies that has examined selection on a reaction norm in a natural population. *Eurosta solidaginis* attacks the goldenrod *Solidago altissima*, laying its eggs in the stems and the larvae causing the formation of a gall. From field data, Weis and

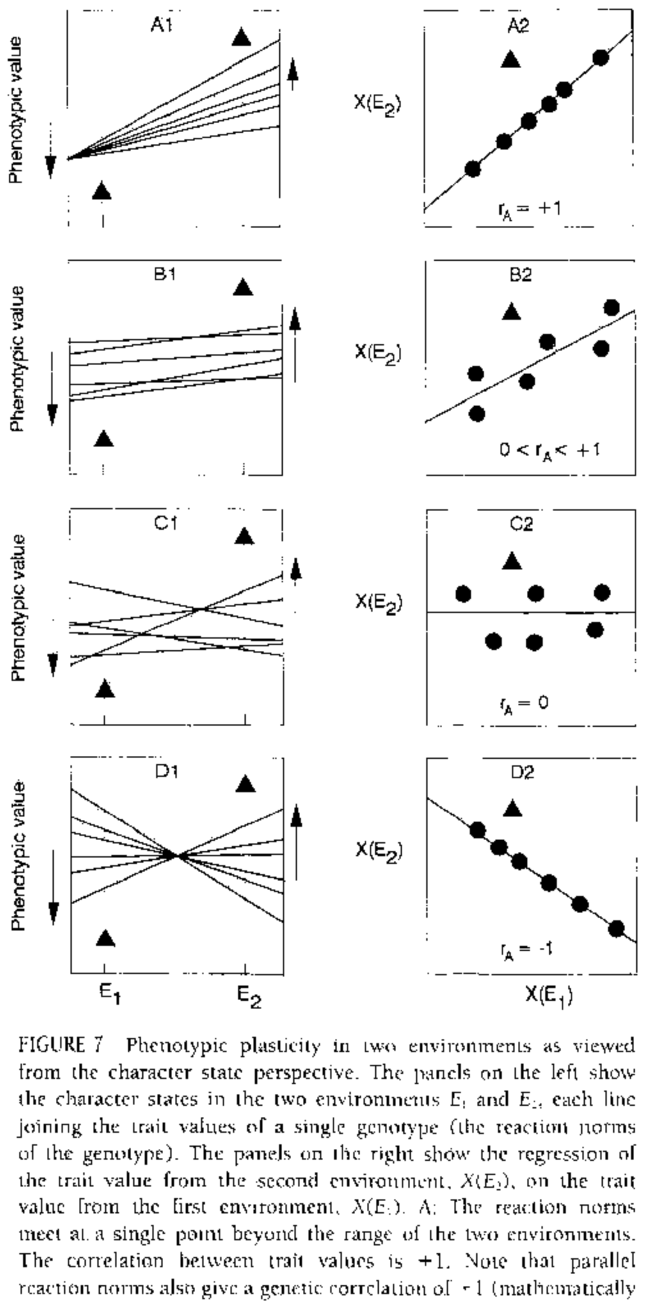


FIGURE 7 Phenotypic plasticity in two environments as viewed from the character state perspective. The panels on the left show the character states in the two environments E_1 and E_2, each line joining the trait values of a single genotype (the reaction norms of the genotype). The panels on the right show the regression of the trait value from the second environment, $X(E_2)$, on the trait value from the first environment, $X(E_1)$. A: The reaction norms meet at a single point beyond the range of the two environments. The correlation between trait values is +1. Note that parallel reaction norms also give a genetic correlation of +1 (mathematically this is because the lines meet at infinity). B and C: The reaction norms cross at several points within the range of the two environments. Depending on the distribution of intersections the genetic correlation will be positive but less than +1 (B), zero (C), or negative but greater than −1 (not shown). D: The norms of reaction intersect at a single point between the two environments. In this case the genetic correlation is −1. The triangles show a hypothetical optimal combination of trait values. Because in cases A and D all the points lie on a single line ($r = \pm 1$), this combination may not be achievable. In all other cases, because in principle the distribution about the regression line is normal (i.e., no value is excluded), selection can move the population to the joint optimum.

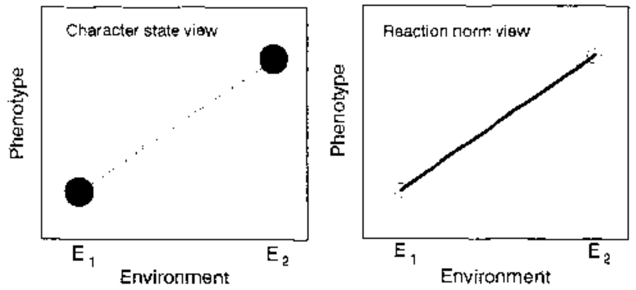


FIGURE 8 A pictorial representation of the two viewpoints of phenotypic plasticity. The character state approach (upper panel) focuses attention on the trait values in the two environments, while the reaction norm approach (bottom panel) focuses on the line joining the trait values.

Gorman determined that survival was a complex function of gall diameter, with an optimum at approximately 24 mm (Fig. 9). On the basis of analysis of 16 full-sib families, they obtained a linear relationship between final gall diameter and the time lag between oviposition and gall initiation (Fig. 9). Because the time lag showed variation associated with plant sibship but not among insect sibship, Weis and Gorman argued that time lag is a trait of the plants and not the insects. Thus time lag is, from the insect's perspective, an environmental gradient associated with the plant. The heritability of the gall-size/lag time intercept was determined as 0.21 (SE = 0.18) and the slope as 0.54 (SE = 0.25). Although the former estimate is not significantly different from zero, the ANOVA indicated significant family effects.

Maximum survival occurs when gall diameter is approximately 24.3 mm, and hence we would expect selection to favor reaction norms that produced a gall of this size, regardless of lag time. In this regard it is significant that the greatest intersection of the reaction norms occurs at the modal lag time but gives a gall diameter of approximately 19 mm (Fig. 9). Thus the fitnesses of the different families are approximately equal in the most frequent environment but are lower than the maximum fitness. Selection should act to shift this area of intersection upward, which could be done by increasing the intercepts or changing the slopes. In the former case an overall increase of approximately 4 mm would achieve the required maximization of fitness within the modal environment, but in the latter case it appears that a much greater range in slopes must be achieved. Therefore, we might expect that selection will act most strongly in a

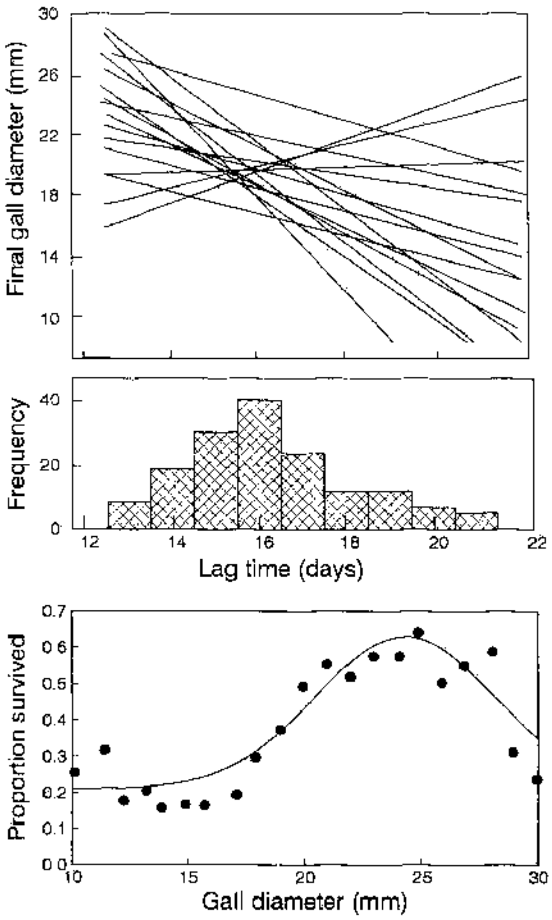


FIGURE 9 Bottom panel: Proportion of survivors as a function of gall size (x) in *Eurosta solidaginis*. Dots show observed value, the solid line the fitted curve

$$\textit{Survival} = 0.21 + 0.42e^{-\left(\frac{x-24.3}{3.84}\right)^2}$$

Middle panel: Frequency distribution of lag time between oviposition and gall initiation. Top panel: Reaction norms from the 16 full-sib families.

directional sense on the intercept, since the same change in all families is likely to have a greater impact on fitness than a similar change in the slope. This prediction can be tested by estimating the relative strengths of selection on the two parameters. As predicted, directional selection was stronger on the slope than the intercept (Intercept/Slope = 4.4), but stabilizing selection acted approximately equally on the two components (Intercept/Slope = 1.4).

This study reinforces the point that the presence of genetic variation is necessary for reaction norms to evolve to their optima, but at the same time the continued presence of genetic variation remains a phenomenon that itself must be explained.

D. Environments with Both Stochastic and Predictable Components

The most realistic scenario of the environment is that it has both stochastic and predictable elements. Consider, for example, the problem faced by a organism in which its propagules (eggs or seeds) can survive frost, and so can overwinter, but all other stages are killed by a frost. Now suppose the season available for growth and reproduction is long enough to accommodate on average two generations. Individuals that mature early in the season will lay direct developing eggs that give rise to the second generation. Individuals of the second generation will all lay diapausing eggs that will overwinter. However, individuals of the first generation that emerge late in the season face the problem that there may be insufficient time for a second generation in which case they should lay diapausing eggs. If the season is long enough to accommodate a second generation, diapause eggs have a lower fitness than direct developing eggs, because the former give rise to only a single descendent for that year whereas the latter give rise to many descendants. In an environment in which the day of the first frost always occurred on the same day or could be predicted without error by the first generation females, there would be a sharp switch between the production of diapause and direct-developing eggs. On the other hand, in general, the day of the first frost could only be assigned as a probability function: under this circumstance there will exist a period during the season during which the most fit strategy will be to lay a mixture of diapausing and direct-developing eggs, the proportion depending on the reliability of the cue.

The phenomenon in which a mixture of types is produced is termed bet hedging or risk spreading. It has been explored theoretically fairly extensively, but there have been few empirical tests. In particular, risk spreading could be achieved by a strategy of avoiding risks (e.g., switching to production of diapausing eggs whenever the probability of not getting in the second generation is less than 1), which has been termed conservative risk spreading or by a single genotype producing a range of phenotypes, which has been termed diversified risk spreading. In an extensive review of insect data Hopper (1999) concluded that there is little evidence that risk spreading has been a major factor in the evolution of insect life histories. The importance of risk spreading in the evolution of life histories remains to be satisfactorily investigated.

See Also the Following Articles

BIODIVERSITY, EVOLUTION AND • BIODIVERSITY GENERATION, OVERVIEW • PHENOTYPE, A HISTORICAL PERSPECTIVE

Bibliography

Boyce, M. S. (1988). *Evolution of Life Histories of Mammals.* Yale University Press, New Haven.

Charlesworth, B. (1994). *Evolution in Age Structured Populations.* Cambridge University Press, Cambridge.

Clutton-Brock, T. H. (Ed.) (1988). *Reproductive Success.* University of Chicago Press, Chicago.

Futuyma, D. J. (1998). *Evolutionary Biology.* Sinauer Associates, Sunderland, MA.

Hopper, K. R. (1999). Risk spreading and bet-hedging in insect population biology. *Annual Review of Entomology* 44, 535–560

Roff, D. A. (1992). *The Evolution of Life Histories: Theory and Analysis.* Chapman and Hall, New York.

Roff, D. A. (1997). *Evolutionary Quantitative Genetics.* Chapman and Hall, New York.

Stearns, S. C. (1992). *The Evolution of Life Histories.* Oxford University Press, New York.

LIMITS TO BIODIVERSITY (SPECIES PACKING)

L. B. Slobodkin
State University of New York-Stony Brook

GLOSSARY

chemostat Apparatus for growing microorganisms in a continually replenished medium.

community A multispecific aggregation of organisms in a particular location that may interact with each other.

competition More than one species utilizing one or more common resources.

ecological niche The set of requirements that must be met if a particular species is to survive. It is sometimes used to mean the place in which those requirements are met.

ecosystem A region with more or less clear boundaries, which contains a particular set of species and may be characterized by some set of meteorological, climatological, and geochemical properties.

invasive or alien species Species that have recently colonized some geographic regions different from the one in which they were initially described.

isocline A line along which some property remains constant.

multidimensional niche A tempero-spatial region defined by meeting a set of different requirements for viability of a particular kind of organism.

niche dimensions Ranges of values of environmental measurements in an ecological niche. A range of temperatures, salinities, or oxygen concentrations may be niche dimensions for a population of fish.

packing The placement of objects in a container.

reification The assignment of empirical reality to the referenda of a word or theory, regardless of the existence of any such referenda.

species diversity The number of different species in some area of interest. Global species diversity refers to all species. Local species diversity refers to some geographic region such as Hawaii or New York City.

species packing The study of how species on the same trophic level coexist in a limited region or container.

SPECIES PACKING is the study of the co-occurrence of different species of organisms in a particular sampling area. Certain regions are known to contain many species of particular groups. Samples of coral reefs, the middle depth sea bottoms, and some tropical forests contain many species while samples from arctic lakes, northern forests, and estuarine bottoms have relatively few. Since it is known that several species simultaneously introduced into a container cannot persist, there is a question about how so many species can be packed into certain situations and not into others. The theory of species packing was a response to this problem.

I. INTRODUCTION: THE BIOLOGICAL PACKING METAPHOR

Different locations vary conspicuously in the number of species that are present. Some, like coral reefs and tropical rain forests, are famous for their high species diversity, while others, like some Alaskan forests, are notably monotonous in the number of kinds of trees that are present. Differences in diversity have attracted many researchers, although as yet there is no one all-encompassing theory of species diversity. In fact, the overall term "species diversity" seems more useful at the level of scientific policy and public relations than it is at the level of scientific discourse.

There are many theories and models of species diversity, which differ in their foci and assumptions. The theory of species packing is concerned with one aspect of diversity, specifically the coexistence of populations of different species located in a particular space. It carries elements of the concept of "species richness" in that the greater the species richness of a region, the more species have presumably been "packed" into it. It also carries implications for "realized species niches" because it might be inferred that the more species are packed into a region, the narrower the realized niches.

Like so many of the terms of theoretical ecology, "species packing" often is used in a metaphorical sense, which can be misleading. It is best to begin by analyzing the concept in a literal and elementary way.

The word "packing" has at least two meanings, which can cause confusion. It may mean placing objects inside a wrapper or container without altering them, or it may mean squeezing them in to a fixed-size container. Contrast the packing of glass tumblers with that of pillows. Tumblers are packed into a container but if they press hard against each other they shatter. Pillows can be squeezed into a container and fluffed up again when they are taken out.

The possibility of either type of species packing presupposes some sort of container into which things are packed. Is the inferred container equivalent to an environment? Does an environment have fixed geometric dimensions like a wooden box? A pond, a log, a rotting fruit—all can, in fact, be thought of as containers. However, environments can be defined in ways that may have more meaning to investigators than to inhabitants. Are the boundaries of a study area like La Selva, a political unit like Costa Rica, or a geographic area like Central America in some sense meaningful packing containers?

We will introduce the mathematical theory of species packing, examine how the theory has been exemplified in laboratory and field experiments, and, finally, consider the implications of species packing for field observations.

II. TWO SPECIES IN ONE CONTAINER

A. Theory

The problem of species packing was crystallized in the apparently simple question: "Why are there so many kinds of animals?" Thereafter there has been a continual stream of often controversial, difficult, and sometimes acrimonious papers on this subject.

1. Lotka-Volterra Equations

Species packing theories are concerned with species in an implicitly assumed *container*. The boundaries of the container are usually considered to be permeable to the flow of resources and energy but, unless specified, impervious to the passage of organisms. These are important assumptions.

The formal theory of species packing originated from the equation systems of Lotka, Volterra, and D'Ancona. Lotka considered that in the entire biosphere all individual organisms have similar resource requirements, and no resource is in infinite supply. Therefore, at some level there must ultimately exist a limit to the abundance of organisms and an inverse relation between number of species and number of individuals per species. Further analysis is required before this assertion becomes interesting.

Volterra and D'Ancona considered several species competing with each other for resources in a space through which resources flowed at a constant rate. Competition between species A and B can be considered in terms of pairs of isoclines on a two-dimensional phase space with axes N_A and N_B:

$$\frac{dN_A}{dt} = 0 = K_A - N_A - \alpha N_B$$

$$\frac{dN_B}{dt} = 0 = K_B - N_B - \beta N_A$$

in which the subscripts refer to the two species. The K values refer to the number of organisms present at what was called "population equilibrium" or "population saturation."

It was assumed that any single species population would remain numerically constant after K had been achieved, if the flow of resources and the physical conditions did not change. These simple equations are the starting point for many species packing theories.

The K and N values were initially considered as counts of organisms, but as the theory developed the K values were taken to designate some resource divided up among the organisms. It often helps to think of species abundance as being directly translatable to rate of consumption of a temporally renewable resource.

It was usually assumed that any individual organism will negatively affect the well-being of all the others in the designated space, as measured by their birth and death rates.

Organisms can affect each other by reducing the availability of resources. They can also alter the chemical or physical properties of the local environment. These alterations are referred to as "conditioning" or "crowding." Yeast cells secrete alcohol, which affects other yeast and bacteria. Beetles and mice change the odor of the environment. Mussels and barnacles physically crowd each other.

In elementary packing theory, a constant, α, represents the relative amount of the resources of species A consumed by each individual of species B and, conversely, β represents the crowding of the individuals of species B by one individual of species A. In the multispecies case, α_{ij} is taken as the effect of species j on members of species i.

2. The Ecological Niche

The possible outcomes for two species were often represented graphically in elementary texts by using a phase diagram with abundance or concentration of each of the two species as axes. The persistence of both species populations is only possible if the two isoclines cross.

For species N_A, the isocline connects the point $(N_A = K_A, N_B = 0)$ to the point $(N_A = 0, N_B = K_B/\alpha)$. At each point on this isocline, individuals of the relevant species are crowded enough that births and deaths are equal. The crowding is exercised by either con-specifics, the members of the other species, or some mixture of the two.

If two species are inoculated into a space in which each species can survive if alone, there are basically three possible outcomes:

- If an individual of one species impacts the con-specifics more than hetero-specifics, both species may persist. The two isocline cross at a "knot."
- Another possibility is that individuals of both species are more sensitive to the presence of the other species than to the presence of con-specifics. In this case, the outcome depended on the initial concentrations of the two species. The isoclines cross at a "saddle."
- The final case is one in which one species has a stronger effect on the other, but the second species does not have a very strong effect on the first. The isoclines do not cross.

These three cases are related to simple definitions of the ecological niche. In the first case, the container is considered to contain regions from two ecological niches. In the second case, the container is in the intersection between two niches, and in the third case the container is within the niche of the victorious species.

B. Laboratory Experiments

While the elementary theory is primarily of pedagogical rather than practical value, it did inspire several laboratory population studies in which populations of two species were actually placed in a container, provided with renewable resources, and permitted to interact.

The early investigations on macroscopic organisms required long periods of observation and repeated tedious counting. Some later laboratory population experiments were done. Current studies of laboratory population dynamics mainly use microorganisms and are primarily of genetic and evolutionary significance. Even in these studies there are significant ecological differences between results derived from liquid culture in test tubes, in chemostats, and those from colonies on semisolid gels.

Without attempting to review all laboratory studies it is possible to summarize insights that have proved relevant in designing field studies and that have suggested theoretical analyses or may be expected to do so in the future.

- Laboratory studies have demonstrated that a single container can be divided in ways that can be more interesting than simply differences in quantitative utilization of resources.

For example, when populations of a moth and of a beetle were placed in a container with intact wheat grains or with flour, the moths were eliminated. Both species survived in cracked wheat. Broken bits of glass capillary tubing added to flour also permitted two species persistence. The moth larvae were being eaten by the beetles unless they were sheltered inside the grains or in the glass tubes. This illustrated that spatial complexity can modify the relation between competing populations. In the absence of physical refuges, the relationship between the moth and the beetle becomes a predator-prey interaction combined with some competition for food. Examples in which the interaction between several species can vary between competition and predation are not uncommon.

No populations, except for those of some microorganisms, grow according to the Verhulst Pearl logistic equation of sigmoid growth. This is interesting because there still are those that will defend the use of this counter-to-fact equation in development of theory.

- As populations become more crowded there are changes in birth and death rates. These changes are always accompanied by other physiological, anatomical, and behavioral changes.

There is abundant evidence of psychological and physiological change caused by individual social history in mammals. While the effects of past history may be most dramatic in mammals, they have been found in essentially all animals and plants. Crowded trees show different shapes than those grown in uncrowded situations.

- The levels of population density that can be reached during population growth experiments are very much in excess of what the organisms would tolerate in nature, as indicated by the ubiquitous tendency of animals to escape from these containers in any way possible.

In one study it was possible to compare responses of house mice in closed and open spaces. When the confined mice became sufficiently numerous to empty their food trays they immediately developed reproductive and behavioral pathologies. Unconfined mice remained in a local area around their food trays until they could empty them after each feeding. They then immediately dispersed.

Flour beetles when crowded move to the surface of their medium and show a strong tendency to fly. This cannot be demonstrated in population vials but becomes apparent if the flour containing a crowded population of beetles is removed from the vials or the vials are unstopped (personal observation). Dispersal when crowded is also found in other grain pests.

In Daphnia, winter eggs (ephippia) are produced at an early stage of food stress. This permits descendants of the population to survive despite whatever occurs in the experimental container. Any pond or lake that contains Daphnia or other Cladocerans will contain abundant ephippia in its sediments. These may hatch after a long dormant period, introducing genetically distinct animals from an animal seed bank for many years, even if the emergent animals of a particular genotype cannot survive in most years.

Even Hydra, which have no locomotory organs, respond to crowding and food shortage by floating free of their substrate, which removes them from a locally crowded situation.

- Laboratory populations can sometimes be used to make some inferences about what may be occurring in nature, but the inferences are not necessarily direct.

Early students of Daphnia assumed that Daphnia *condition* the water in which they lived. It was later demonstrated that the sizes of Daphnia populations in the laboratory, over a crowding range 10 times that ever observed in nature, were linearly dependent on food supply. If conditioning was of importance, this linearity would be impossible and there would have been a curvilinear relation similar to that found in the relation between nutrient flow and population size in chemostat cultures of microorganisms.

This does not imply that Daphnia populations in nature are regulated by their food supply. It does imply that they are not regulated in any significant way by causing chemical deterioration of their environment, because over a range of crowding that greatly exceeds any that has been found in nature, no chemical effect has been found.

Clearly, laboratory experiments and theoretical models cannot adequately imitate nature. However, an experiment can be designed to consider a wider range of conditions than those that occur in nature. In that sense, nature becomes a subset of the experimental world.

In some cases there is a clear logical transition between theory, laboratory experiment, and field experimentation. Gause mathematically considered predation on a two-species competition system. He concluded that the coexistence of two species could, in principle, be stabilized by predation.

In laboratory populations, the coexistence of brown and green species of hydra was facilitated by either removal of some of the animals or by maintaining them in low illumination, demonstrating that coexistence could be stabilized by either a biological or a physical change in the container.

In a field experiment, Paine found that heavy predation of starfish on mussels prevented the mussels from eliminating their competitors. In this case, theory proposed a possibility, laboratory experiments demonstrated its reality, and the field experiment demonstrated its importance in nature. But even in this case there were dangers of misinterpretation. The theory and the laboratory and field experiments demonstrated that under some circumstances a predator might enhance the number of surviving species on the next lower trophic level.

Any loss of organisms imposed on a multispecies system relieves, to some extent, competitive pressure on the survivors, so that weaker competitors may survive. There is, however, an abundance of cases of predators locally eliminating prey species, as is now occurring due to feral cats in Australia and Hawaii.

III. EXPANDING THE THEORY OF PACKING

Many formal theories of species packing begin by complicating this two-species situation. Among the obvious complications are those in which the isoclines are permitted to be curved lines, which may be interpreted as social interactions of various sorts. Also, changing the sign of the interactions could crudely mimic predation or symbiosis. Loss of organisms across the "container" boundary could be thought of as mimicking predation.

A. Expansion of the Theory to Multiple Species

The more general question is: How many species can be introduced into a space? Multiple species packing theory is mathematically more difficult because a multispecies theory cannot be conveniently represented as isoclines on a two-dimensional phase space. If a third species were to be added, the phase space would have to be three-dimensional and isoplanes would be needed instead of isoclines. For coexistence of three competing species, the three isoplanes would have to intersect somewhere in the region below the intersection of any two of the species isoclines.

In the case of multiple species packing, there can be a region in the multidimensional phase space in which resources are divided in such a way that all species can share, but only if each species is relatively strongly self-limiting and relatively mild in its effects on the other species present. The multiple species theory requires clarification of the notion of *ecological niche*. The ecological niche of a species is often considered as the full set of measurements of environmental properties that are relevant to the survival and persistence of that population. Hutchinson and his students developed this into an image of a multidimensional hyperspace.

If two hyperspaces do not contain common dimensions, then the occurrences of the species concerned are independent of each other. As more and more dimensions are identical, or very similar, the stronger are the competitive interactions between the two species. Competition is one partial explanation for replacement of similar species along gradients in such physical features as temperature and salinity.

B. How Similar Can Coexisting Species Be?

MacArthur and Levins asked: What might be expected if a third species invades a space in which two species are already resident? They simplified the situation by considering that all species concerned interact through competition on a single niche dimension, or at least a linear representation of a multidimensional niche. They also assumed simple distribution of resources along that dimension. If, for example, the niche dimension is some particular prey, one of the species might eat large specimens while the other eats only small ones.

They considered that the niche of any species would have a shape on any single niche dimension. If the niche can be represented by a straight line, the probability of an individual finding a bit of resource at some point on the line can be represented as an ordinate, generating a two-dimensional shape. If the shapes for several species overlap, then the intensity of competition is measurable in terms of the probability of the several species striving for the same resource—the product of the individual species-specific probabilities.

They could then consider species packing along a niche dimension. Given their assumptions, packing can be closer on any one dimension as the number of niche dimensions increases. If species are arranged in a one-dimensional array along a line, the closest stable packing is admissible when the niche spaces are rectangular and all K values are equal. These conclusions were strongly dependent on their assumptions.

More general conclusions were that closer packing is possible if niche dimensionality is high, niche breadth is small, and the environment is predictable and has high productivity. Centers of adjacent niches are in a multiplicative series of numerical value approximately 1.1.

If the single linear dimension of competition is considered more realistically, so that there is a multiplicity of niche dimensions, MacArthur and Levins concluded that interspecific competition could be less important. In fact, it is sometimes considered that multispecies packing involves as many niche dimensions as there are species. Tilman considers that this is only valid if the organisms are not fixed spatially (i.e., they are free to move about in the container).

Various authors, by slightly modifying the mathematical assumptions, could reach a variety of conclusions about the packing of niches. For example, it was possible to plausibly argue that niches could sometimes be packed infinitely closely.

MacArthur and Levins suggested, as a matter of convenience, that the shapes of niches projected onto a single niche dimension might be thought of as approximating normal distributions. Other shapes are, of course, imaginable.

Species packing became a rich field for construction of theories, which were not necessarily connected to specific biological observations and more. These theories each make slightly modified assumptions or focus on special cases of previous theories. For example, Roughgarden and Feldman considered the significance of shape differences as if the shapes could be empirically demonstrated. Distributions with "thick" tails permitted more species to coexist than those with thinner tails.

There were also studies of multispecies biological situations under natural or artificial circumstances, which used the term "species packing." In these the connection with species packing theory was sometimes explicit but very often the general conclusions of theory were presented in the context of field data without explicit theoretical derivation. This appears to be due to the fact that explicit formal theory production becomes very difficult once more realistic elements of natural complexity are involved. This is typical of mathematical models, at least prior to the recent development of powerful computing machinery.

Sometimes the models were sufficiently complex, combining ecological and evolutionary predictions, that self-contradictory assertions were made within the same theoretical argument and were not noticed for years. For example, one study published in 1983 assumed absence of any genetic heterogeneity and then proceeded to discuss the genetic effects of selection, as noted by Taper and Case.

C. Niche Separation and Anatomical Difference

The exact ratio of minimum possible distinction between adjacent species can be derived from special, simplified, and somewhat arbitrary assumptions. Depending on assumptions, it can be asserted that an infinite number of species can coexist, that an infinite number of species could coexist were it not for statistical variance in the parameters, that coexistence is contingent on number of dimensions in the hyperspace, or that species must differ anatomically by a specific ratio in order to coexist.

Hutchinson suggested the possibility that one might find in nature that niche hyper-volumes of the most closely similar, often congeneric, coexistent species would actually differ from each other by a factor whose value might be approximately 1.3. This seemed reasonable based on the common observation that genera are often represented in particular locations by large, small, and medium-sized species. Attention was focused on this by Lack for Darwin's finches, but its occurrence can be easily confirmed by examining intertidal snails of the American northeast coast or the eastern shores of the Mediterranean Sea (personal observation).

"Assembly rules" for competing species on islands were developed in part on the basis of size. These results were criticized on statistical grounds. Were the species distributions really departures from random?

One major step was missing in attempting to demonstrate that a theoretical niche overlap in a simplified mathematical model predicted a morphological ratio between anatomical parts. While Hutchinson expected that trophic structures like jaws might show the appropriate tightly packed ratio, the theoretical connection between niche shape or size and body shape or size seems absent except in very special cases.

It was never really clear why there should be a relation between the overlap of niche hyperspaces and any obvious anatomical differences. There is no rigorous and general way of theoretically deciding which of the infinite possible number of anatomical measurements that might be made is appropriate in any given case. By choosing the measurements correctly or by manipulating the data, almost any ratio might be achieved. Simberloff provided a firm critique of measuring the niche separation by anatomical differences. There is a clear connection between body size and the utilization of preexisting nest holes by snakes, owls, chipmunks, weasels, and others, but any pair of nest hole utilizing species is almost certain to differ in many niche dimensions.

IV. SPECIES DIVERSITY AND SPECIES PACKING

Does species packing set limits to species diversity in nature? This would require that natural aggregations of organisms are organized into communities and their component guilds, which would act as "containers." Are natural aggregations of organisms in nature strongly mutually interdependent so that "packing in" more species or removing species already present should have discernible effects on those already present? Do cohesive communities of organisms or ecosystems actually exist in nature? Early ecosystem theorists believed in a "holistic" or emergent concept of ecosystems and communities. If these were real, they would constitute containers in which packing theory might be applicable.

There is an enormous literature related to how many species occur in particular locations. Recently there has been a weakening of the basic notion of a natural community as a "container" in any serious sense. There is massive concern about the dangers of invasive species, but recent studies have returned emphasis to individual species and almost discarded the idea of integrated communities. There is a general condemnation of invasive species but they are particularly interesting from the standpoint of species packing. If invasions do not occur, it may be due to lack of opportunity. Large mammals cannot float across an ocean or logs or hide away in ballast tanks. However, if a species can invade a region, then in some sense the region was not packed full, or perhaps successful invasions necessarily result displacements of resident species.

The apparent invader purple loosestrife is sometimes found in dense stands that keep out other species, but other studies find that purple loosestrife had no apparent effect on native plant species and may have benefited native insect diversity.

Ruiz estimated that more than 90% of alien species in estuaries have made no discernible impact on the species diversity or species abundance distribution of the estuary. Levine and D'Antonio report that a "consistent positive relation between exotic species abundance and resident species diversity . . . [suggesting] that invaders and resident species are more similar than often believed." Apparently some invasive species fit into niches that were in some sense empty.

To what degree is competition, and therefore species packing theory, a serious factor in nature and how could one tell? Attempts to answer this question sometimes hinged on naturalists' insights, sometimes on various statistical analyses, and sometimes on detailed natural history and experiments conducted in more or less natural circumstances. Each of these has its strengths, advocates, and opponents.

In what sense do communities depend for their continuity on particular species? American chestnut trees constituted almost a third of the large trees of the southeastern American forests in the early 19th century. While small specimens still exist, the chestnut trees were essentially wiped out by the Chestnut blight. The forest did not disappear. As noted by Hairston *et al.*, the southeastern forest is as dense with trees as it has ever been. The general appearance of the forests have not changed, but the oaks, which have to a large degree replaced the chestnuts, produce volatile organic compounds that are of biogeochemical significance (M. Lerdau, personal communication, 2000).

Davis, studying the northward expansion of forests after glacial retreat, noted that each individual species seemed to migrate at its own rate—there was no movement of the forest as a community marching together. Whittaker noted that distribution of trees on various environmental gradients was as individual trees, not as communities.

V. CONCLUSIONS

Theory, experiments, and natural history all suggest that communities are not tightly organized, so that species packing may not be strongly relevant in nature. In short, the popular image of an ecological community as an airplane in which each part has a vital role for the integrity of the whole is dubious. There may be groups of species, in which each one is closely connected to a few others but only loosely connected to other groups.

The term "community" was once extremely useful and is still of pedagogic value if carefully used. As reified objects for research, the concept of communities is now threatening to become what Simberloff and Dayan have referred to as a "panchreston," an idea as likely to generate confusion as enlightenment.

The prospect of a general species packing theory has melted away. Despite the unreality of the models, they did direct attention to what seemed to be a real phenomenon and encouraged experimentalists and field biologists to ask important questions.

Differences in species richness can be partially explained by a multiplicity of factors that do not necessarily relate to each other so as to permit formation of a single coherent theory. Each of the multiplicity of theories of diversity focuses on one or a few of the empirical factors that are known to enhance or diminish the possibility of species coexistence.

Many of these can be seen in laboratory experiments, which have the advantage of clarity but may be of questionable applicability. For example, local geometry complexity influences species diversity in nature as well as in the laboratory.

Certainly in some cases the term "species packing" is used in the sense of species being squeezed into a space. In several cases individuals of particular species in species-rich regions are believed to have a narrower range of activities than individuals of the same species in species-poor situations. Diet or nest sites may be more restricted. In these cases the individual organisms can be imagined to have been constricted by some packing process, like individual pillows in a crate but even this has two possible meanings depending on whether we are concerned with the population level or with individuals.

In comparing different locales, the range of variation among organisms within a unispecific natural population may be reduced when more other species are present. In the case of comparisons of different islands or lakes, this might be tentatively attributed to species packing.

Returning to the initial analogy of packing actual objects, an island is clearly a container. But if packing means filling the ecological space, either by pillows or tumblers, we would not expect islands or speciose lakes to easily admit invasive species. In lakes there may be enormous species richness of fishes, as in the ancient African lake cichlids, but I don't know of any comparisons showing that species-rich lakes show less within species, among individual variation, than species-poor lakes. Is the attempt to crowd multispecific collections of fish together in the same lake equivalent to packing glass tumblers rather than pillows? There is a general impression that species rich systems seem at least as likely to be invaded by exotic species as species-poor systems, violating our sense of what packing might mean.

If the container walls are not apparent but among individuals variation is reduced in one or more populations of a species, is this a sign of species packing? Does the mere fact that among individuals variation is reduced when more species are present imply that there must exist a container wall which may not be obvious?

The overall conclusion is that the theory of species packing does not conveniently predict very much about natural systems but that the images of packing that it generates do informally suggest interesting phenomena to look for.

See Also the Following Articles

COMPETITION, INTERSPECIFIC • DIVERSITY, COMMUNITY/REGIONAL LEVEL • HABITAT AND NICHE, CONCEPT OF • PLANT INVASIONS • SPECIES-AREA RELATIONSHIPS • SPECIES COEXISTENCE • STRESS, ENVIRONMENTAL

Bibliography

Abrams, P., Nyblade, C., and Sheldon, S. (1986). Resource partitioning and competition for shells in a sub-tidal hermit-crab species assemblage. *Oecologia* 69, 429–445.

Adams, D. C., di Bitetti, M. S., Janson, C. H., Slobodkin, L. B., and Valenzuala, N. (1997). An "audience effect" for ecological terminology: Use and misuse of jargon. *Oikos* 80, 632–636.

Anderson, M. G. (1995). Interactions between Lythrum salicaria and native organisms: A critical review. *Environmental Management* 19, 225–231.

Ayala, F. (1971). Competition between species: Frequency dependence. *Science* 171, 820–824.

Barrett, G. W., Peles, J. D., and Odum, E. P. (1997). Transcending processes and the levels-of-organization concept. *Bioscience* 47, 531–535.

Bazzaz, F., Ceballos, G., Davis, M., Dirzo, R., Ehrlich, P. R., Eisner, T., Levin, S., Lawton, J. H., Lubchenco, J., Matson, P. A., Mooney, H. A., Raven, P. H., Roughgarden, J. E., Sarukhan, J., Tilman, G. D., Vitousek, P., Wall, D. H., Wilson, E. O., and Woodwell, G. M. (1998). Ecological science and the human predicament. *Science* **282**, 879.

Berryman, A. A. (1992). Intuition and the logistic equation. *Trends in Ecology and Evolution* 7, 316.

Boake, C. R. B., and Wade, M. J. (1984). Populations of the Red Flour Beetle Tribolium castaneum (Coleoptera, Tenebrionidae) Differ in Their Sensitivity to Aggregation Pheromones. *Environmental Entomology* 13, 1182–1185.

Burgman, M., and Lindenmayer, D. (1998). *Conservation Biology for the Australian Environment*. Surrey Beatty and Sons, Chipping Norton, NSW.

Cornell, H., and Lawton, J. (1992). Species interactions, local and regional processes, and limits to the richness of ecological communities: A theoretical perspective. *Journal of Animal Ecology* 61, 1–12.

Crombie, A. (1945). On competition between different species of graminivorous insects. *Proceedings of the Royal Society of London Biological Sciences* 132, 362–395.

Crombie, A. (1946). Further experiments on insect competition. *Proceedings of the Royal Philosophical Society of London, Series 3* 133, 76–109.

Daily, G. C., Alexander, S., Ehrlich, P., Goulder, L., Lubchenco, J., Matson, P., Mooney, H., Postel, S., Schneider, S., Tilman, D., and Woodwell, G. (1997). Ecosystem Services: Benefits Supplied to Human Societies by Natural Ecosystems. Ecological Society of America, Washington, D.C.

D'Ancona, U. (1954). *The Struggle for Existence*. E. J. Brill, Leiden.

Davis, M. (1986). Climatic instability, time lags, and community disequilibrium. In *Community Ecology* (J. Diamond and T. Case, Eds.) pp. 269–284. Harper and Row, New York.

Diamond, J., and Case, T. (1986). *Community Ecology*. Harper and Row, New York.

Dykhuizen, D. (1978). Selection for tryptophan auxotrophs of Escherichia-coli in glucose-limited chemostats as a test of energy-conservation hypothesis of evolution. *Evolution* 32, 125–150.

Elton, C. (1958). *The Ecology of Invasions by Animals and Plants*. Methuen and Co., London.

Feldman, M. W., and Roughgarden, J. (1975). Populations stationary distribution and chance of extinction in a stochastic environment with remarks on theory of species packing. *Theoretical Population Biology* 7, 197–207.

Ferson, S., Stewart, S., Downey, P., Jacques, G., Klerks, P., Ssemakula, J., Weissburg, M., Malenky, R., Kroot, L., and Anderson, K. (1986). Competing reviews, or why do Connell and Schoener disagree? *American Naturalist* 127, 571–576.

Fryer, G., and Iles, T. D. (1972). *The Cichlid Fishes of the Great Lakes of Africa: Their Biology and Evolution*. Crown, New York.

Gause, G. F. (1934). *The Struggle for Existence*. Williams and Wilkins, Baltimore.

Glasser, J. W. (1983). Variation in niche breadth with trophic position—On the disparity between expected and observed species packing. *American Naturalist* 122, 542–548.

Griffing, T. C. (1965). *Dynamics and Energetics of Populations of Brown Hydra*. Ph.D. Thesis, University of Michigan.

Hairston, N., Sr. (1991). *Ecological Experiments: Purpose, Design and Execution*. Cambridge University Press, Cambridge.

Hairston, N. J., Lampert, W., Caceres, C., CL, H., LJ, W., Gaedke, U., Fischer, J., Fox, J., and Post, D. (1999). Rapid evolution revealed by dormant eggs. *Nature* 401, 446.

Hairston, N., Smith, F., and Slobodkin, L. (1960). Community structure, population control and competition. *American Naturalist* 94, 421–425.

Harvey, P., Colwell, R., Silvertown, J., and May, R. (1983). Null models in ecology. *Annual Review of Ecology and Systematics* 14, 189–211.

Hopf, F. A., and Hopf, F. W. (1985). The role of the Allee Effect in species packing. *Theoretical Population Biology* 27, 27–50.

Hutchinson, G. E. (1959). Homage to Santa Rosalia or Why are there so many kinds of animals? *American Naturalist* 93, 145–159.

Hutchinson, G. E. (1978). *An Introduction to Population Biology*. Yale, New Haven.

Kinzig, A. P., Levin, S. A., Dushoff, J., and Pacala, S. (1999). Limiting similarity, species packing, and system stability for hierarchical competition–colonization models. *American Naturalist* 153, 371–383.

Lack, D. (1947). *Darwin's Finches: An Essay on the General Biological Theory of Evolution*. Cambridge University Press, Cambridge.

Leopold, A. (1966). *A Sand County Almanac: With Other Essays on Conservation from Round River*. Oxford University Press, New York.

Levine, J., and D'Antonio, C. (1999). Elton revisited: A review of evidence linking diversity and invasibility. *Oikos* 87, 15–26.

Lomnicki, A., and Slobodkin, L. B. (1966). Floating in hydra littoralis. *Ecology* 47, 881–889.

Lotka, A. J. (1925). *Elements of Physical Biology*. Williams and Wilkins, Baltimore.

Lotka, A. J. (1934). Theorie analytique des associations biologiques. *Actualites Scientifique et Industrielles* 187, 1–45.

MacArthur, R. (1972). *Geographical Ecology*. Harper and Row, New York.

MacArthur, R., and Levins, R. (1967). The limiting similarity, convergence, and divergence of coexisting species. *American Naturalist* 101, 377–386.

Mahmood, T., Ahmad, M. S., and Ahmad, H. (1996). Dispersion of stored grain insect pests in a wheat-filled silo. *International Journal of Pest Management* 42, 321–324.

May, R. (1972). *Model Ecosystems*. Princeton University Press, Princeton.

Mullin, B. (1998). The biology and management of purple loosestrife (Lythrum saliceria). *Weed Technology* 12, 397–401.

Odum, E. P. (1992). Great ideas in ecology for the 1990s. *Bioscience* 42, 542–545.

Paine, R. T. (1966). Food web diversity and species diversity. *American Naturalist* 100, 65–75.

Pantasticocaldas, M., Duncan, K. E., Istock, C. A., and Bell, J. A. (1992). Population-dynamics of bacteriophage and Bacillus-Subtilis in soil. *Ecology* 73, 1888–1902.

Pianka, E. (1975). Niche relations of desert lizards. In (M. Cody and J. Diamond, Eds.), *Ecology and Evolution of Communities*. pp. 292–314. Harvard University Press, Cambridge.

Pimm, S. (1991). *The Balance of Nature?: Ecological Issues in the Conservation of Species and Communities*. University of Chicago Press, Chicago.

Ritte, U. (1969). *Floating and Sexuality in Laboratory Populations of Hydra Littoralis*. Ph.D. thesis, University of Michigan.

Roughgarden, J. (1974). Species packing and competition function with illustrations from coral-reef fish. *Theoretical Population Biology* 5, 163–186.

Roughgarden, J., and Feldman, M. (1975). Species packing and predation pressure. *Ecology* 56, 489–492.

Ruiz, G. M., and Fofonoff, P. (1999). Non-indigenous species as stressors in estuarine and marine communities: Assessing invasion impacts and interactions. *Limnology and Oceanography* 44, 950–972.

Schoener, T. W. (1985). Some comments on Connell and my reviews of field experiments on interspecific competition. *American Naturalist* 125, 730–740.

Simberloff, D., and Boecklen, W. (1981). Santa Rosalia reconsidered: Size ratios and competition. *Evolution* 35, 1206–1228.

Simberloff, D., and Dayan, T. (1991). The guild concept and the structure of ecological communities. *Annual Review of Ecology and Systematics* 22, 115–143.

Slobodkin, L. (1954). Population dynamics in Daphnia obtusa Kurz. *Ecological Monographs* 24, 69–88.

Slobodkin, L. (1961). *Growth and Regulation of Animal Populations*. Holt, Rinehart and Winston, New York.

Slobodkin, L. (1964). Experimental populations of Hydrida. *British Ecological Society Jubilee Symposium. Journal of Animal Ecology* 33(Supplement), 131–148.

Slobodkin, L. (1992). *Simplicity and Complexity in Games of the Intellect*. Harvard University Press, Cambridge.

Slobodkin, L. (1993). Scientific goals require literal empirical assumptions. *Ecological Applications* 3(4), 571–573.

Strecker, R., and Emlen, J. (1953). Regulatory mechanisms in house-mouse populations: The effect of limited food supply on a confined population. *Ecology* 34, 375–385.

Taper, M. L., and Case, T. J. (1992). Coevolution among competitors. In *Oxford Surveys in Evolutionary Biology* (D. Futuyma and J. Antonovics, Eds., vol. 8, pp. 63–109. Oxford University Press, Oxford.

Tilman, D. (1994). Competition and biodiversity in spatially structured habitats. *Ecology* 75, 2–16.

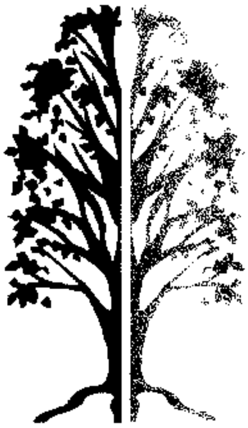
Treberg, M., and Husband, B. (1999). Relationship between the abundance of Lythrum salicaria (purple loosestrife) and plant species richness along the Bar River, Canada. *Wetlands* **19**, 118–125.

Volterra, V. (1926). Variazione e fluttuazione del numero d'individui in specie animali conviventi. *Mem. Accad. Naz. Lincei* **2**, 31–113.

Whittaker, R. (1970). *Communities and Ecosystems*. MacMillan, New York.

Yan, G. Y., Stevens, L., Goodnight, C. J., and Schall, J. J. (1998). Effects of a tapeworm parasite on the competition of Tribolium beetles. *Ecology* 79, 1093–1103.

Yoshiyama, R. M., and Roughgarden, J. (1977). Species packing in 2 dimensions. *American Naturalist* **111**, 107–121.



LITERARY PERSPECTIVES ON BIODIVERSITY

William Howarth
Princeton University

GLOSSARY

contingency Possibility or uncertainty; an event that may occur but is not likely.

ecocriticism Interdisciplinary study of literature, history, religion, and philosophy with an emphasis on places, evolutionary biology, and environmental problems.

etiology The study of causes or origins; stories that explain the origins of phenomena.

land ethic Valuing land as part of a biotic community, not merely as property.

literature Imaginative and crafted writings, in the form of poetry, prose, fiction, or drama.

nature The material world and the physical forces or processes that control it.

pastoralism Agrarian life and work; also literary accounts of rural life, often simplified or idealized.

places Areas of space with human association; also positions of status in society.

"DIVERSITY" REFERS TO DIFFERENCES or variations; "biodiversity" refers to the variety among biological species, sometimes called species richness (Nudds, 1999). Biodiversity is a recent concept, used since the 1980s to publicize losses in wildlife habitat, genetic variety, and number of species (Wilson, 1992). Field and laboratory studies confirm that natural diversity sustained a healthy planet until human beings began to acquire global dominance through technology and urbanization. Recent policies of sustainable development, especially in first-world nations, seek to restore biodiversity with land-use planning, protection of endangered species, and preservation of wild habitats. Although most definitions of biodiversity focus on its economic value, something beyond cash receipts must account for the continuing interest of societies in nature, wildlife, and open space (Werner, 1990).

The idea of biodiversity traces back to early concepts of variety and unity, which have their sources in history, philosophy, and literature. These cultural fields have recognized the need for biological variety since the first days of recorded time. Literature comprises many acts of language, from ancient folk chants to poetry on the Internet, often for entertainment and instruction, the *dulce et utile* praised by Cicero in his writings on rhetoric. The main goal of literature is to expand and extend

Encyclopedia of Biodiversity, Volume 3

the human imagination. By simulating experiences and performing ideas, literature helps affirm the value of diverse places and species, through a wide range of styles and themes. Not all writings are considered literature, just as all experiments are not science. Like professional science, the ideal of literature implies standards of custom and cultural decree (Turner, 1985).

I. LITERATURE AND SCIENCE RELATIONS

Over the centuries, literature has come to mean imaginative writing, the product of disciplined craft that creates drama, fiction, poetry, or essay. Critical theorists disagree as to whether the prime agent in literature is content or form, conveying a message or shaping an attractive verbal structure, like a story or a poem. In the balanced formula of British poet and critic Samuel Coleridge, poetry is "the best words in the best order," which affirms that an ideal of quality guides literary writers. But "quality" is an intrinsic value, one that clashes with the extrinsic aims of science. Science seeks an unbiased view of what nature is and how it operates, while literature observes those conditions through the filters of human belief and emotion.

Although novelist C. P. Snow (1959) believed that science and literature are opposing cultures, recent observers find more congruency in their aims. Both disciplines use observational methods to detect patterns and to record them for the benefit of posterity. Biologist Stephen Jay Gould (1989) sees science and literature as forms of historical narrative that describe contingency, the recording of uncertain, unexpected events just beyond current knowledge. Literature uses contingency in plot sequences—as events connect and readers turn pages "to find out what happens next." Scientists unravel nature's "plots" by reconstructing them as predictive narratives.

Literary forms represent natural diversity with varying degrees of accuracy. Drama is imprecise because it relies on social situations and voiced sentiments, enacted on a physical stage. Painting and photography may document a place or species, but pictures without words are not sufficiently interpretive. Poetry (whether epic or couplet) and the various forms of prose (novel, story, essay, memoir) are considered more comprehensive because they both describe and evaluate, mingling sensory experience with calculated thought. The meaningful connection of image and word: literature pursues that goal in order to represent biodiversity for the benefit of human understanding.

II. LANGUAGE IN SCIENCE AND LITERATURE

Human beings share a common attribute in language, the instrument they use to construct meanings. The path of language moves from variety to unity: many words exist; when combined they form phrases that express thoughts. A thought may be social or personal, as it creates a send-and-receive cycle that communicates. To meet the human need to understand both surface and depth, literature often asserts that natural objects are not alien but connected to human emotions. As Henry David Thoreau wrote in *Walden* (1854), "Shall I not have intelligence with the world? Am I not vegetable mould myself?" Seeking knowledge of nature's variety is thus a mutual goal for literature and science, but sharp differences exist in their methods.

Scientists prefer to use words precisely, to convey strictly limited meanings. Literary writers recognize that words have long histories, in which meanings change across time and space. As philologists learned in the 19th century, words behave like organisms (growing from root to stem to branch; cross-breeding and migrating to other habitats). Some words are viruses, infecting a host, while others lie in dead heaps, like sediment. Through these morphological changes, words shape human values to form and reinforce perceptions, beliefs, and opinions.

Words are also representational, using speech sounds to create signs for rain, snow, leaf, and rock. Combinations with other words produce symbolic and metaphorical overtones, suggesting layers of possible meaning. Ezra Pound's haiku-like poem, "In a Station of the Metro" (1916), compresses a startling comparison of culture and nature into two lines: "The apparition of these faces in the crowd;/Petals on a wet, black bough." Such powers of analogy give words a suggestive and speculative function, which lends itself to stories with a didactic purpose.

III. ETIOLOGICAL TALES OF NATURE

The earliest surviving fables and myths are etiological tales, told to explain how the world began, why rain falls, and when crops will bloom or die. Among prehistoric people, nature stories about diverse and complex phenomena became a form of cosmology, providing the basis for ritual ceremony, whether for medical or spiritual healing. Early ideas of disease, as a condition of imbalance or disharmony, encouraged the symbolic belief that the earth was also a vast "body" that may be cleansed and healed. Some observers (Buell, 1998) note that medicinal rhetoric continues to shape current discourse on toxicity, pollution, plague, and other forms of ecodisaster.

Early tales are efforts to define the myriad natural forms that shape everyday life, from land and species to climate and seasons. At the same time, literature may portray human efforts to resist or control nature, commonly through acts of labor and art. Farmers drain wetlands in order to plant crops on dry land. They weave nets to lift fish from the sea, then bury the fish with seeds to help new life emerge. Stories about these actions form early contributions to understanding biodiversity by recognizing the complexity of nature and the intricacy of its related elements. In a Navajo creation tale, "Changing Woman and Her Hero Twins," the storyteller pours tinted sands on the earth, forming a series of objects and events that are as multiple and shape shifting as the identity of Changing Woman, an earth goddess who has the power to create.

Yet by emphasizing human impact, many early stories also teach that humans and nature are divisible and separate. That idea is distinctly Judeo-Christian and supports the very concept of "nature," meaning all that is born, rising from a source. No equivalent term appears in Asian culture, where the physical world connects humanity to the wider universe (Torrance, 1998). The Old Testament tells two different stories of human creation: man is made separately, in the image of God (Genesis 1:26), but is also formed from dust (Genesis 2:7). The first story appears to justify human dominion over other creatures, while the second suggests a humbler bond with the earth. Belief in a separate, special creation of human beings continues to guide modern fundamentalists, who fiercely resist Charles Darwin's unitary theory of evolution. By including two narratives, the Bible leaves open the problem of literal versus metaphorical interpretation.

IV. PASTORALISM IN POETRY AND SCRIPTURE

In the early civilized era, 4000 to 500 B.C., ideas and tales of nature's variety emerged from the extractive activities of hunting, fishing, farming, herding, logging, and mining. By learning to harvest natural materials and process them into goods, humans became experienced observers of seasonal events, of plant and animal lives, and of locations favorable to cultivation. Land use and property rights became the basis for literacy, trade, law, and tribal identity (Crumley, 1994). Nations root themselves in physical places, and when people are forcibly displaced, as in the case of the early Jews, they endure their nomadic years with consoling legends of a Messiah, a warrior-shepherd who will return them to the lost homeland. Such early stories reveal a basic function of literature, to reflect the value of places and how human presence changes them.

In the Old Testament, land is an element made by God, then granted to humans to clear and own. They receive this gift in innocence, but later sin and must accept the burdens of mortality: labor, pain, disease, and death (Genesis 1-3). The status of wilderness is even less benign: uncultivated places, like mountains and deserts, display God's power and offer a test to the faithful. In the story of Noah (Genesis 6-9), God directs Noah to build a great Ark and bring to it at least two of every creature that lives, male and female. Then God floods the earth with rain, until only the Ark and its inhabitants survive. The story illustrates not just divine power, but also how human stewardship may preserve the earth's vast diversity of species. The later vision of Isaiah predicts a savior who will restore the earth to its unfallen state, when all species live together in harmony: "The wolf also shall dwell with the lamb, and the leopard shall lie down with the kid; and the calf and the young lion and the fatling together; and a little child shall lead them" (Isaiah 11:6). This scene, oft painted by the 19th-century Quaker artist Edward Hicks, represents the differences between a creation marred by original sin and one healed by the world's innate bounty.

In times of constant social change, literature offered comforting accounts of bygone days and places. The very concepts of city and country, or urban and rural land, arose out of literature that invested both realms with a complex relation, possessing zones and boundaries of separation but also connection (Williams, 1974). As urban centers grew into the great city-states of Alexandria, Athens, and Rome, the poets Hesiod, Theocritus, and Virgil wrote pastoral poetry, a literary form known variously as bucolic, georgic, and eclogue; verses that portray rural life as gentle, principled, and close to the soil. Jesus of Nazareth also adapted pastoralism in his parables about fish and seeds to preach transcendence of earthly limitations, a literary device equally effective with audiences in rural Judea and in the fast-growing cities of the Mediterranean basin.

After the Roman empire shifted from paganism to Christianity in the third century A.D., early Christians saw their pastoral mission as spreading the Gospels throughout Europe. Monastic enclaves preserved both classical science and literacy in the Dark Ages (400–800), between Rome's fall and the revival of learning in courts and cathedral towns. During the Middle Ages (800–1400), infectious plagues reduced populations and relieved pressure on Mediterranean ecosystems, still recovering from Roman-era deforestation, erosion, and pollution (Hughes, 1994). Medieval literature and science became twinned endeavors, as clerics and scholars sought to describe the fixed Ptolemaic universe, an effort that culminates in Dante's epic poem, *The Divine Comedy* (1321). Medieval culture's fascination with alchemy likewise fed early studies of natural history and chemistry, described in several of Chaucer's *Canterbury Tales* (1400).

V. COLONIAL EXPANSION AND TAXONOMY

Chaucer's use of folk tales reflects the popular literacy that spread rapidly after Johannes Gutenberg invented movable type (1455), as printed tracts, manuals, and libraries of books fed a new hunger for learning. Literacy spawned the Renaissance (1500–1700), a period of European expansion into the continents of Africa, Asia, and the Americas. Global exploration refined mapping and navigation, while imperial conquest decimated native populations through warfare and disease. Discoveries of new lands and species also spurred growth in the descriptive sciences, especially geography, geology, and biology. A buoyant verbal fluency stirred the language arts, epitomized in the writings of Shakespeare, Cervantes, and Molière. Colonial empires used their literacy to record events and claim land as property. Native cultures often regarded words as totemic: their oral chants and stories were prayers, repeated to bless earthly cycles. Such radically different uses of language also expressed conflicting views of nature. Europeans sought to own the earth, even as their sciences and arts revealed the immense biodiversity of wild, untouched land.

With the rise of natural science after 1700, literature began to express growing respect for nature, especially those aspects of it that were rare and varied. Optical devices peered into the heavens and the human body, while theorists and engineers created new branches of knowledge, publicized and marketed by print, commerce, and coffeehouse (Jardine, 1999). New forms of technology changed the economic value of natural resources. Land remained a source of food and energy, but wealth increasingly came from products and services. Cities expanded by drawing rural tenants from the countryside; new social classes formed around labor, management, and capital. An expanding, literate middle class called for political reforms, replacing monarchy with representative democracy. Revolutions in the Americas pulled masses of immigrants to the new Western nations, where open territory and cheap land began to transform agrarian life into republican independence. (As the world has grown tamer in recent decades, wilderness has come to have a higher value, as land existing for itself, rather than for human enterprise.)

The study of natural diversity advanced through what Thomas Kuhn (1967) has called changing paradigms, theories or conventions that promoted experimentation. Early naturalists described organisms with confusing and inconsistent local vernacular until Carolus Linnaeus compiled *Systema Naturae* (1734), using Latin nomenclature to assign names by genus and species. Latin was dead and thus fixed, and Linnaeus used social distinctions (kingdom, phylum, class, order, family) to create taxonomy, a systematic language that described the bounty and variety of nature. While literary authors from Jonathan Swift to Fenimore Cooper mocked scientists for obsessive categorizing, taxonomy became a consistent model for geology, archaeology, and philology. Those sciences declared that prehistoric changes in the earth and in language follow traceable lines of descent across long stretches of time. Such discoveries replaced mythic accounts of creation with rational studies of the unity and diversity of organic events.

VI. DARWINISM AND LITERATURE

The new sciences spawned a second age of exploration, 1750 to 1850, as voyagers from James Cook to Alexander von Humboldt surveyed the continents of Australia and the Americas. In their wake followed a generation of literary naturalists like Gilbert White, William Bartram, and John James Audubon, who drew field sketches and wrote evocative accounts of seasonal journeys through rural or remote lands. The expeditionary narratives of Peter Kalm or Meriwether Lewis and George Rogers Clark are read today as literature and history, for they verify the immense variety of species that led to theories of biodiversity. This tradition of gentleman amateur molded Charles Darwin, who spent a 5-year expedition on the *HMS Beagle*, examining patterns in geology and zoology along the coasts of South America and on Pacific islands. Darwin was one of the last major scientists to be an independent field naturalist, living on his own income. After 1870 science moved to the laboratory bench and university departments, a shift in professionalism that also affected writers, who became teachers and seekers of literary prizes (Jardine, 1996).

In his major works, *Voyage of the HMS Beagle* (1839), *The Origin of Species* (1859), and *The Descent of Man* (1871), Darwin argued that species evolve through the effects of heredity, variation, and natural selection. He had several predecessors in science, from Georges-Louis Leclerc Buffon, who recognized adaptation, to Alfred Russell Wallace, who anticipated and clarified several of Darwin's ideas about selection. The term "evolution" is quite ancient, going back to Latin *evolvere*, to unroll, and may describe inscriptive scrolls, the writings that physically expressed a gradual process of change. The Biblical account of creation (Genesis 1) frames a sequence of originating events, each a division (night from day, land from water, plant from animal) that measures the binary logic of doubling, branching, and splitting, all recurrent patterns in nature. Poetry, tragedy, and biography all depict similar sequences of choice and accident, selection and reproduction. Those lines of cause and effect link the Bible to Homer and Sophocles, and then to Shakespeare or Milton. A constant theme in Western literature is the relation of creator, man, and nature, as chains of interactive events lead to human survival or extinction.

Literary traditions emerged from worship and theology, while Darwin stated that evolution proceeds according to physical laws, without divine intervention, and that human beings did not arise separately from other species. Early confirmation of these views appeared in the research of the Austrian monk Gregor Mendel, whose experiments in hybridization found that reproduction is a process of encoding and transmitting genetic material. Despite the strong controversy these views aroused among people of faith, by the 1860s evolution and genetic science began to alter popular ideas of nature and produce the first authors who consciously recognized biodiversity.

Two leaders in this movement were Henry David Thoreau (1817–1862) and George Perkins Marsh (1801–1882). Their books (*Walden, Cape Cod, The Maine Woods, Man and Nature*) often examine the physical environment of New England, a region of low economic value but rich biodiversity. A self-trained land surveyor, Thoreau spent several years studying landforms and seasons. After reading *The Origin of Species* in 1860, he wrote the first modern accounts of forest succession and the dispersal of seeds (Howarth, 1982). Marsh, a lawyer and diplomat, published early studies of erosion, blaming land clearance and overgrazing, and he argued vigorously that human beings must learn to control their destructive impact on nature.

Among later writers known as social Darwinists (chiefly Herbert Spencer, Walter Bagehot, and William Graham Sumner), application of evolutionary theory to human behavior produced racist theories of cultural history, favoring white and Nordic peoples over Mediterranean and African. This bias, based on fears of immigration by darker races toward the wealthier nations, slants the novels of Jack London, Frank Norris, and Theodore Dreiser. Later these voices were replaced by modernist writers like T. S. Eliot, Willa Cather, and Ernest Hemingway, who saw racial diversity as healthy and progressive, a human check on industrial monopolies that had lain waste to natural resources and open spaces. Cather explored these issues on the plains of Nebraska, while Hemingway examined waning biodiversity in several lands, from upper Michigan to the savanna of east Africa, where great forests and wild herds are steadily destroyed by human incursion.

VII. ECOLOGY AND INDUSTRIALISM

The concepts of ecology were slow to gain acceptance among biologists and thus entered literary culture at a later date. Ernst Haeckel, a German biologist and advocate of Darwin, coined the term "ecology" in 1869 to describe how organisms form alliances that shape their number and distribution. Based in part on Haeckel's political socialism, his views also echoed rising concern that industrial growth in Germany, France, and England had begun to damage both physical environments and public health (Bowler, 1994).

A vivid picture of decline lay in the western United States, where rapid agrarian settlement after the Civil War destroyed native tribes, bison, and grasslands in a few decades. The conservation movement of 1870 to 1920 sought to stem those losses by creating public lands and parks for protection of natural resources. Yet in the popular arts, writers of dime novels and outlaw ballads glorified "winning the West," a form of historic triumphalism sanctioned by the theories of Frederick Jackson Turner (1893), who held that the frontier shaped fundamental American values of self-reliance and innovation. Later scholars note that the settlement process was destructive, as pioneers destroyed the very Eden they had sought (Mitchell, 1981).

With the experimental writings of James Joyce, Gertrude Stein, and William Faulkner, a modern critique of mass-industrial civilization began to emerge. Drawing on their perspectives as cultural outcasts (Irish Catholic, Jewish Lesbian, Southern Gentry), they mounted strong attacks on the spiritual vacuity of urban life and, in Faulkner's case, on the economic abuses that had sacked agrarian America. In his 14-novel saga about a fictional county in northern Mississippi, Faulkner assailed profiteering, an ideology that created slavery, miscegenation, one-crop farming, and soil erosion. He was the first modern novelist to comprehend ecological concepts, which gained attention in the 1930s from New Deal efforts to counter the wide-scale failure of capitalism.

The word "ecology" connotes a desire to coexist, if not in symbiosis, at least in a mix of competition and collaboration that marks the dynamics of a healthy, vigorous ecosystem. Darwin's image of nature was "an entangled bank," and throughout the 20th century, ecologists created an interdisciplinary science that depicted nature as a web of interacting, interdependent forces that sustains itself—not as an unchanging constant, but as a vigorous dynamism sustained by frequent, patchy disturbances (Hagen, 1992).

Three writers who affirmed these principles, in sharply varying styles, were all Californians: poet Robinson Jeffers, "hard-boiled" novelist Raymond Chandler, and John Steinbeck, a social novelist who was trained in marine biology and in the investigative methods of documentary journalism. In his literary works, especially *The Grapes of Wrath* (1939) and *The Sea of Cortez* (1941), Steinbeck built a vision of interlacing natural and cultural forces, locked in struggles that were shaped by biological conditions of production, consumption, and adaptation (Beegle, 1997).

VIII. LAND ETHICS AND ENVIRONMENTALISM

After World War II, two writers emerged who were trained ecologists and willing to apply their influence to public policy. In their writings, Aldo Leopold and Rachel Carson issued firm warnings of environmental destruction by human means. Leopold's *The Sand County Almanac* (1949) celebrates a glaciated wetlands in Wisconsin, poor in economic value but rich in its diversity of habitats. His farm lies at the intersection of prairie, forest, and marsh, forming a mosaic of ecosystems that measure the "intricate tangle" of nature. Out of this vision, Leopold sketches the story of how he abandoned early conservationist ideas (a wildlife ranger, he was paid to kill wolves) for a preservationist vision he calls "the land ethic," the sense that land deserves respectful protection as a vital element in the biotic community.

In *Silent Spring* (1962), biologist Rachel Carson intensified that sense of respect by analyzing the destructive force of chemical pesticides. In an incisive passage, she demonstrates how toxicity spreads through groundwater, a system of transport so invisible that it is easy to ignore. For her efforts, Carson was attacked by the petrochemical industry as hysterical and unscientific. But her logic and eloquence impressed the Kennedy administration, brought a ban on DDT and other pesticides, and eventually helped create the Environmental Protection Agency (Howarth, 1999).

Both Leopold and Carson saw that human values are formed by ideas of land, seeing that property is also a shared earth. Their books brought attention to emerging "green" political movements of the late 20th century, from which two different land ethics developed: landscape, shaped by people and inflected by aesthetic beauty, and land use, focused on food and energy extraction and marked by utilitarian security. Although often in conflict, the two ethics also reflect agreement that human identity rests on a sense of place, while alienation is feeling displaced, homeless, or unlanded (Spirn, 1999).

IX. NATURAL RIGHTS AND ANIMAL FABLES

The perception that land deserves ethical regard soon led to extending rights to nonhuman species, both plant and animal, granting them value and respect while seeking to conserve their populations. Herman Melville's epic novel *Moby-Dick* (1851) anticipates this generosity, expressed in an iconoclastic rhetoric that denounces whaling. The cause of animal rights began with 19th-century reformers, who sought protection for religious dissenters, children, women, Indians, slaves, and other powerless groups. These figures came to be seen not as "lower" orders, but creatures having what Darwin called "the mutual affinities of organic beings." That phrase, the title of his penultimate chapter in *Origin of Species*, delivers his belief that humans are descended from all the creatures that have existed and still share the earth.

This theme enters literature in many animal tales written for young readers, from Anna Sewell's *Black Beauty* (1887) to E. B. White's *Stuart Little* (1945) and

Charlotte's Web (1952). They descend from ancient beast and fairy tales, in which animals have the power to create language, before passing it on to humans. Such stories describe how things came into being, and they still have great attraction for writers drawn to environmental themes (Flynn, 1999). A contemporary example is David Quinn's novel *Ishmael* (1991), a series of dialogues between man and ape, in which the ape explains history as a conflict between Takers and Leavers, identified by their divergent approaches to the natural world. Takers dominate resources and evolve into modern urbanized man; Leavers are indigenous tribes who lead subsistence lives and eventually vanish.

While such accounts are fantastic, they also urge readers to think of animals, water, soil, rocks, and food as more than commodities. The cultural information provided in literary texts clarifies that natural objects result from processes, including the moral and ethical choices made by humans. As such choices recur in history, the value of a creature or a place changes: once the sea was thought to contain monsters, later it became a therapeutic place to bathe and sun; now it is emptying of fish and clogged with sewage (Hendrickson, 1984). Formerly the forest world was seen as profane and lawless, the home of pagans; now it is sacred, a haven for rest and retreat (Harrison, 1992).

X. ECOCRITICISM AND BIODIVERSITY

Scientists are now warning that a global environmental crisis will be the future outcome of several current trends: human overpopulation, the rise of carbon dioxide emissions, atmospheric warming, and the loss of biodiversity. These dangers transpire across vast frames of time and space that are difficult for nonscientists to imagine. A crisis that occurs in a locality, such as flood or earthquake, is understandable, but one that works over a longer period and through evolutionary selection is invisible to most eyes.

Because its domain is virtual or imaginative, literature helps its readers to envision what they will never directly experience. Yet many variables also affect the ability of literature to depict the threat to biodiversity. A writer's personal involvement in a story will shape its tone, bringing nostalgia (Wallace Stegner) or urgency (Bill McKibben) to the subject matter. The work's form may illustrate its content, as when Peter Matthiessen shapes *The Snow Leopard* (1973) into a mountaineering ascent, using elevation to provoke meditations on higher metaphysical themes. An object described will take on the character of its subject, as when Annie Dillard expresses her notions of wildness through a weasel's alleged indifference to pain.

The work may also have a distinct cultural audience (John McPhee) or veer between standards of elite and popular reception (Edward Abbey). Some writers are sensitive to social and economic conditions (Edward Hoagland), others contrast regional and national concerns (Barry Lopez), still others attend to differences of race and gender (Sue Hubbell), or to the impact of other arts on literature (Anne Matthews). Environmental writing is often a mixed collection of genres; part scholarly research, part sermon-editorial, part whimsy and adventure. Similar manifestations affect the contemporary novels of Thomas Pynchon (*Gravity's Rainbow, Vineland, Mason and Dixon*) and Don DeLillo (*White Noise*).

Because of the great variation in environmental writing, a new mode of reading called "ecocriticism" has emerged to establish interpretive standards. While not agreeing in all respects, ecocritics often seek to translate the specialized language of science, with its jargon and acronyms, into a common vernacular. Ecocritics are interpreters and bridge builders, looking for ways to connect the disciplines that university departments have long separated (Glotfelty and Fromm, 1995). A few scientists, such as Lewis Thomas (1974) and E. O. Wilson (1998), have written popular accounts of their environmental concerns, but mostly the public has turned to literary writers to gain insight into the causes and effects of waning biodiversity. At times these voices cannot agree about solutions, for the global biosphere is too intricate, too complex, and too unpredictable to describe readily, and some scientists believe that it will not be managed until it returns to pre-industrial levels of stable composition (Peters, 1994).

Other ecologists have grave concerns about genetic erosion, through both species loss and also genetic engineering, which in the long term may produce fewer varieties of plants and animals. (Karaim, 1999). The role of species biodiversity suggests that its purpose may be to provide redundancy, which has a stabilizing effect on whole communities and ecosystems. In the face of drastic environmental change, diversity may buffer ecosystems against the collapse of ecological function. (Nudds, 1999).

If biodiversity works because variety protects a community against stress and disaster, then literary accounts of biodiversity must remain equally diverse and inventive for the sake of textual survival. A review of literary history suggests that the continuance of human life on earth will depend in part on what storytellers make of a changing world. The gift of language brings insight, but also the wisdom to heed the silence that Annie Dillard calls "nature's one remark." The central question for human beings is not whether they will survive as a species, but whether the survivors will inherit a world worth sharing (Wilson, 1998).

See Also the Following Articles

AESTHETIC FACTORS • DARWIN, CHARLES • HISTORICAL AWARENESS OF BIODIVERSITY • RELIGIOUS TRADITIONS AND BIODIVERSITY • SOCIAL AND CULTURAL FACTORS • STEWARDSHIP, CONCEPT OF

Bibliography

Beegle, S. (1997). *Steinbeck and the Environment: Interdisciplinary Approaches*. University of Alabama Press, Tuscaloosa.

Bowler, P. (1994). *The Norton History of the Environmental Sciences*. W. W. Norton, New York.

Buell, L. (1998). Toxic discourse. *Critical Inquiry* 24, 639–665.

Crumley, C. L. (1994). *Historical Ecology: Cultural Knowledge and Changing Landscapes*. American Research Institute, Santa Fe, NM.

Flynn, K. M. (1999). *Peaceable Kingdoms: Constructions of Animal Life in American Literature*, 1850–1950. Ph.D. Dissertation, Princeton University.

Glotfelty, C., and Fromm H. (1995). *The Ecocriticism Reader: Landmarks in Literary Ecology*. University of Georgia Press, Athens, GA.

Gould, S. J. (1989). *Wonderful Life: The Burgess Shale and the Nature of History*. W. W. Norton, New York.

Hagen, J. (1992). *An Entangled Bank: The Origins of Ecosystem Ecology*. Rutgers University Press, New Brunswick, NJ.

Harrison, R. P. (1992). *Forests: The Shadow of Civilization*. University of Chicago Press, Chicago.

Hendrickson, R. (1984). *The Ocean Almanac*. Doubleday, New York.

Howarth, W. (1982). *The Book of Concord: Thoreau's Life as a Writer*. Viking Press, New York.

Howarth, William (1999). Imagined territory: The writing of wetlands. *New Literary History* 30, 509–39.

Hughes, D. (1994). *Pan's Travail: Environmental Problems of the Ancient Greeks and Romans*. Johns Hopkins University Press, Baltimore.

Jardine, L. (1996). *Cultures of Natural History*. Cambridge University Press, Cambridge.

Jardine, L. (1999). *Ingenious Pursuits: Building the Scientific Revolution*. Little, Brown, New York.

Karaim, R. (1999). Variety, the vanishing crop. *Washington Post*, April 11, B01.

Kuhn, T. S. (1967). *The Structure of Scientific Revolutions*. University of Chicago Press, Chicago.

Mitchell, L. C. (1981) *Witnesses to a Vanishing America: The Nineteenth-Century Response*. Princeton University Press, Princeton.

Nudds, T. (1999). Harvest In *Status and Trends of the Nation's Biological Resources*. NSGS Web Site: http://biology.usgs.gov/s+t/SNT/noframe/he118.htm.

Peters, R. L. (1994). Global Warming and Biological Diversity. Yale University Press, New Haven, CT.

Snow, C. P. (1959). *The Two Cultures and the Scientific Revolution*. Cambridge University Press, Cambridge.

Spirn, A. W. (1999). *The Language of Landscape*. Yale University Press, New Haven, CT.

Torrance, R. (1998). *Encompassing Nature: A Sourcebook*. Island Press, Washington, D.C.

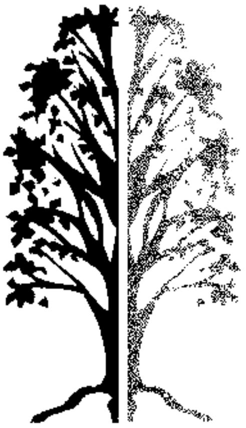
Turner, F. (1985). *Natural Classicism: Essays on Nature and Science*. Paragon House, New York.

Werner, T. B. (1990). *Conserving Biological Diversity*. World Resources Institute, Gland, Switzerland.

Williams, R. (1973). *The Country and the City*. Oxford University Press, New York.

Wilson, E. O. (1992). *The Diversity of Life*. Harvard University Press, Cambridge, MA.

Wilson, E. O. (1998). *Consilience: The Unity of Knowledge*. W. W. Norton, New York.



LOGGED FORESTS

Reinmar Seidler* and Kamaljit S. Bawa*†

**University of Massachusetts, Boston and †Ashoka Trust for Research in Ecology and the Environment, India*

GLOSSARY

alpha diversity The variety of organisms occurring in a particular place or habitat; often called local diversity.

beta diversity The variety of organisms within a region arising from turnover of species between habitats.

canopy Usually the highest tree layer in the vertical stratification of forests. The herbaceous foods for canopy-dwelling animals are found among the branches and leaves of the canopy.

dipterocarp A member of the Dipterocarpaceae family (comprising 22 genera) of South Asian and African timber trees.

ecosystem A community of interacting living organisms and its physical surroundings.

forest management Broadly, a pattern of human activities to derive economic or other utility from a forest. In its narrower sense—as a concept or guiding principle for the multiple and sustainable use of a forest—it is a complex ecological and sociological concept in which exceptional skill has to be exercised on the part of forest manager. Forests could be managed for the extraction of timber or nontimber forest products, forage for animals, watershed protection, or recreational use.

gap An opening in the vegetation created from disturbances such as clearing, logging, fires, diseases, storms or the natural death of a tree.

inbreeding Mating between closely related individuals more often than would be expected by chance.

logging The operation of harvesting trees, sawing them into appropriate lengths (bucking), and transporting them (skidding) to a sawmill. In modern day, this method is mostly mechanized.

mast fruiting A period or year in which a heavy crop of fruits/seeds is synchronously produced by trees and shrubs. This phenomenon, uniquely characterized by synchronicity, high variability, and periodicity of heavy fruit production, distinguishes it from nonmasting plants.

natural regeneration The process of replacement by natural seedlings in gaps created by the selective cutting of marketable timber.

pioneer guild The first groups of species to colonize a newly formed or denuded habitat.

primary forest Forests that appear to be undisturbed by human influence.

secondary forest Successional forests growing in areas where in the past the forest cover had been completely removed.

Encyclopedia of Biodiversity, Volume 3

shifting agriculture The practice of clearing a plot of land for cultivation for a short period of time, then abandoning it and allowing it to revert to its natural vegetation when the cultivation moves to another plot.

silviculture An applied science and branch of forestry concerned with the theory and practice of controlling forest establishment, composition, and growth.

succession Replacement of one kind of population/community by another in a habitat through a regular progression over time.

taungya The Burmese term for an agricultural system in which crops are interplanted with plantations of trees. As the trees grow and shade the areas, cultivation of crops is abandoned.

understory The broad spectrum of plants at the ground level of forests that often provide forage for grazing animals. These transitory plants are often relatively sparse in dense forests with closed canopies.

OVER MILLENNIA, HUMAN SOCIETIES have managed forests for the production of a range of goods and services. As long as population densities remained low, the impact of humans on forest ecosystems was minimal, except near population centers of ancient civilizations. With the start of the industrial revolution, the demand for forest products increased considerably. At the same time, rapidly expanding human populations started to exert tremendous pressure on forests. The demand for timber (and nontimber forest products) grew at an accelerated pace. Timber extraction on an industrial scale became the principal goal of forest resource managers in the latter half of the 19th century and remained so throughout the 20th century. Today, there is virtually no forest on the planet without at least some signs of human influence. With mounting losses of forest cover and increasing interest in the conservation of biodiversity, attention has now again shifted to managing forests for a wider array of goods and services. In this new paradigm shift, management plans may include logging as an element among others, or they may not include logging at all.

I. INTRODUCTION

Clearly, forests can be managed for a variety of goals. The impact of different management systems on biological diversity is still largely unclear, although there have been a number of studies of forest ecosystems managed for the production of timber. Here we provide a perspective on changes in biodiversity in forests managed largely for timber. The concepts and principles of large scale forest management for a diverse range of goods and services are still being developed, even though traditional societies have managed forests for such purposes for centuries. We first outline various systems of management and then describe the extent of our knowledge about their effects on forest biodiversity. The effects of these systems vary according to a number of factors, including the type of forest, the landscape matrix in which the forest is embedded, the nature and intensity of the management system, and the scale and method of analysis chosen by the observer. Forests are complex ecosystems with subtle and complex interrelationships, many of them unknown in detail and changing over time. The potential for variation in each of these factors is great, making generalization across specific cases tenuous.

Forest diversity may be examined at various levels, including ecosystem, landscape, population, and genetic levels. It is advantageous to keep these multiple levels in mind as one considers forest management systems. Since the changes effected by management on forest biodiversity do not happen all at once, but are mediated through a chain or sequence of influences both spatially and over time, our discussion is organized sequentially. We begin by describing the *structural* changes imposed on forests by management operations or activities. These structural changes determine the ensuing *abiotic* or environmental conditions within the forest at the small and medium (local and regional) scales. Structural changes and altered abiotic conditions combine to produce a variety of *biotic* changes, including changes in species composition and abundance, productivity, population density and distribution, and so forth. It is thought that these changes may also affect long-term *ecological* interactions and processes. Thus, the original structural alterations to forests result in changed environmental conditions, which in turn trigger changes at the species, population, and community levels. Large reductions in effective population sizes may result in *genetic* alterations and impoverishment. In addition, it is important to keep in mind that virtually every managed forest experiences a variety of indirect *synergistic* anthropogenic effects, namely human activities not part of the management itself but facilitated by it. These may include increased incidence of wildfire, hunting, shifting agriculture, and others. Taken together, these activities may multiply the overall impact of the management system.

II. MANAGEMENT SYSTEMS

Systems of forest management vary widely throughout the world in response to ecological, social-institutional, and political factors. Use of the phrase *management system* need not imply that management is in each case applied in a planned, thoughtful, or consistent manner, but only that it is systematic in the sense of being a pattern of human manipulative activities governed by a specific common goal. These activities may be carefully planned and controlled interventions or they may be haphazard patterns of one-time individual extractive activities, without much overall guidance. Thus, we use the word *management* rather broadly.

A. Timber Management

1. Clearing Systems

At one extreme of a continuum of management systems for the extraction of timber are systems of forest clearing, with or without replanting to stimulate regeneration. *Clearing systems* have been used most often in temperate forests. These forests have relatively low numbers of tree species and therefore a relatively high density of commercially useful species per hectare. However, some tropical forests, such as some of the dipterocarp-dominated forests of Southeast Asia, may also have sufficiently dense stands of commercial trees to make clear-cutting an economic option. If there are residual noncommercial trees, they may be girdled or poisoned to produce even-aged stands in which commercial species predominate. There have been experiments with various geometries of clear-cuts, intended to aid in natural regeneration and the preservation of wildlife habitat. Examples are the checkerboard patterns in the Pacific Northwest of the United States and Canada and the strip shelter-belt system of Palcazu in Peru. Patterns of this type have as their goal the reduction in area of the individual gaps, and thus the reduction in distance between the edge of the remaining forest and the center of clear-cut areas. The underlying idea is to increase the likelihood of propagules from the forest finding their way into the clear-cut area and regenerating the stand. There have also been attempts to develop timber management systems that use natural gaps as models, but these are considered selective logging systems (discussed below).

Other forest management systems involving clearing are *replacement systems*, in which plantations of indigenous or exotic species are established after clearing. These have been widely used in temperate forests and less so in the tropics, partly because they often involve a larger initial investment in silviculture. The even-aged stands which result are seen as advantageous in terms of accounting and the efficiency of silvicultural operations, but they have been vulnerable to outbreaks of pests and disease and, as highly simplified ecosystems, have been criticized as deleterious to biodiversity.

2. Selective Logging Systems

In many species-diverse tropical forests, clear-cutting is neither an economically nor an ecologically interesting option, since only a few tree species are commercially accepted. Moreover conspecifics are widely scattered, and infrastructure including roads and rural mills may be sparse. In such managed forests, trees are selectively logged. *Selective logging* is a general term encompassing a wide array of management systems that vary widely with respect to spatial and temporal scale, harvesting intensity, planning, and oversight. Selectivity of species to be removed does not necessarily translate into selectivity of overall impact, so the residual forest may be affected indiscriminately even by the selective extraction of only a few trees per hectare.

Tropical forest management systems were adapted from the German forestry tradition, which was exported, to India and Burma during the 19th century. *Natural regeneration systems*, or *polycyclic systems*, minimize silvicultural interventions by relying on natural regeneration after the harvest of a relatively low number of trees per hectare. These systems result in uneven-aged and multispecies stands, which are thought to provide the best opportunities for biodiversity conservation as part of the management plan. However, a variety of assumptions about regeneration patterns must be made, many of which have been questioned by some foresters, and conclusive evidence of long-term sustainability is lacking.

Natural regeneration systems with adaptations to regional conditions have been developed in Malaysia (selection management system), Ghana (modified selection system), Suriname (Celos silvicultural system), Trinidad (periodic block system), and Queensland, Australia (Queensland selective logging system). Each of these has made important strides toward solving technical silvicultural problems, but in many cases sociopolitical obstacles have been more severe.

B. NTFPs

Timber is not the only product to be taken out of forests and marketed, nor is it the only product for which forests are managed. Nontimber forest products

(NTFPs) include fruits, nuts, fungi, fibers, medicinal and ornamental plants, mosses, dyes, resins, gums, fuelwood, charcoal, leaves as fodder, poles for local construction, honey, syrup, fish, and game, as well as other animal products. In some forests, these may constitute important and large-scale commercial resources; in others, they may have great local importance but fail to appear in commercial markets. Management systems for NTFPs, too, run the gamut from traditional, socially sanctioned systems to those that are legally organized and monitored. In some cases, extraction of NTFPs may affect forest biodiversity even more than timber extraction, since NTFP extraction is often done over a long period of time—over many human generations in parts of southern Asia, Southeast Asia, and Africa—and may constitute an intensive use of some species or areas. Most of the economically significant NTFP extraction systems are today found in tropical forests and include extraction of Brazil nuts, rattans, ornamental plants, and animals, fruits, and medicinal plants. Maple syrup is an economically significant and commercially developed NTFP of temperate forests.

C. Shifting Agriculture and Traditional Long-Term Intensive Forest Management Practices

As primarily agricultural systems, the effects of shifting agriculture on biodiversity are treated extensively in another chapter. Shifting agriculture is by no means the only traditional forest management practice, however. Some traditional practices have been very intensive, especially in high population density areas of Asia, Europe, and Central America, and some have been sustained over the very long term. In Japan and Europe, coppicing and pollarding systems, in which trees are repeatedly pruned back for their leaves or poles, have been practiced for centuries. These traditional systems have preserved habitat for a diversity of plants, birds, arthropods, fungi, reptiles, and amphibians that are absent from younger successional forests in the same areas. Traditions of pasturing domestic animals in commons woods, and the retaining of hedgerows, copses, and windbreaks within intensively managed agricultural landscapes, have all contributed to the survival of woodland diversity in Europe (Kirby and Watkins, 1998). In the tropics, traditional systems of *taungya* and other intensive agroforestry systems have for centuries combined high forest productivity with high species diversity. *Taungya* is a Burmese word, but similar traditions of agroforestry and the combination of tree crops with ground crops and kitchen gardens may be found in Brazil, Indonesia, Malaysia, and the Indian subcontinent. There are currently attempts to reintroduce and adapt some of these traditional techniques to modern conditions, since they are thought to hold promise both ecologically and economically.

Sacred groves found throughout the world that have been "managed" for centuries for spiritual, religious, and utilitarian reasons are beyond the scope of the present discussion.

D. Managed Forests as Distinct from Secondary Forests

The phrase *secondary forest* has been used in a variety of ways in the scientific literature of the recent decades. We reserve the phrase *secondary forest* for successional forests growing in areas where the forest cover has at some time in the past been completely removed, causing a break in the continuity of the vegetative cover over time, and where this break in continuity can be detected structurally or floristically. Forest historians, geographers, and archeologists have been pushing back the date for earliest detectable human impacts on forests throughout the world, and it is becoming clear that few if any forests are primeval in the literal sense of the word. Nevertheless, we call forests that show no obvious structural or floristic traces of human influence undisturbed primary forests.

From the perspective of forest biodiversity, the distinction between secondary and managed primary forests is important. Secondary forests consist of earlier successional stages; they are dependent on seed dispersal from outside for their regeneration and the continued process of succession. The implications of this for biodiversity will be discussed later.

III. STRUCTURAL ALTERATIONS

Forest structure is the three-dimensional arrangement of trees and other plants, in combination with nonliving spatial elements such as soils, slopes, and hydrology. In short, structure is the physical geography of the forest, considered at a range of spatial scales. Structure includes such characteristics as canopy and understory geometry, continuity or fragmentation of canopy cover, homogeneity or patchiness of species distribution through the landscape, soil structure, and the species composition and age structure of stands. Most of these elements may be considered at micro, local, or landscape scales. Forest structure influences forest biodiversity directly through the formation of microhabitats as well as the determination of larger-scale habitat characteristics, but the relationship between structure and species diversity is complex and not well understood.

The most immediate effects of selective timber extraction systems on forest structure, of course, are the removal of the individual target trees, together with the associated incidental damage. Incidental damage generally includes removal of surrounding vegetation to construct access roads and skid trails, damage to vegetation during felling and skidding, compaction, and scraping of soils. Harvest of as little as 3% of the trees in an area, as is not atypical in the diverse neotropical forests, may reduce canopy cover by half through incidental damage. Canopy reduction by 75% is not uncommon in Asian dipterocarp forests, where a total of nearly half the basal area of the forest may be removed. On the other hand, damage to basal area of as little as 4 or 5% has been documented in Amazonia, when less than a single mahogany tree per hectare was removed. And noncommercial harvesting projects have obtained even better harvest-to-damage ratios in Costa Rica, Ghana, Madagascar, and Queensland, Australia.

In conventional operations, the impact on undergrowth can also be substantial. In a logged area in neotropical French Guyana, nearly half of the undergrowth was destroyed during the removal of an average of only three trees per ha. Removal and damage of vegetation opens the canopy, creating gaps and artificial edges, and often lowers the average height of the canopy, thus altering the internal vertical habitat structure of the remaining forest and, in some cases, removing most or all individuals of the larger and older size classes.

Little precise information exists on the sorts of structural damage ensuing from NTFP extraction. It is clear that damage from intensive harvesting of leaves for fodder, poles for construction, fruits, and firewood can all be substantial.

In the temperate zone forests, the structural changes in managed forests often result from changes in species composition and age structure of overstorey trees, as the multispecies stands are replaced by even aged populations of commercially valuable native or exotic species. Such changes set in motion a number of changes in ecosystem structure (biodiversity) and function.

A. Vertical Structure

The vertical stratification of forest vegetation has been the subject of some debate among plant ecologists. It is clear that many forest organisms (birds, insects, and other arthropods, certain herps) partition habitat in the vertical dimension. It is less clear what effect timber management may have on this dimension of forest structure. Selective timber extraction may simplify structure, for instance, by eliminating or preferentially damaging one or several strata. On the other hand, it has been suggested that one reason for the increase in local abundance of certain species groups in secondary forests is the greater complexity of the understory structure in younger, less well developed forests. The enhanced presence of short-lived treelets, shrubs, and herbaceous plants in disturbed forests is a matter of record (Richards, 1996).

B. Canopy

The forest canopy, the highest layer of the vertical structure, includes most of the interface between leaf and light. As such, it is the area of greatest energy input into the forest ecosystem. In tropical or temperate evergreen rain forest, this interface may be virtually unbroken over large areas, and there is often a very distinct division, in terms of biodiversity, between the canopy layer and the understory. In seasonal, dry, or gallery-type forests, the canopy may be much more permeable to incoming light and correspondingly less distinct in its flora and fauna.

C. Gaps and Edges

The creation of forest gaps through disturbances such as selective logging both enhances and diminishes species diversity and richness in different ways. Gaps increase the local heterogeneity of habitats and create microhabitats that are rare in mature forest, but they simultaneously limit the extent of undisturbed forest, thereby reducing regional habitat heterogeneity. The question of which tendency will predominate is at the heart of discussions about management for biodiversity. The answer depends partly on the characteristics of the focal species or group, partly on the extractive methods employed, and partly on the spatial scale of analysis. *Alpha diversity*, which is a measurement of species diversity at the community level, does not necessarily covary with *beta diversity*, namely the overall turnover of species associations in contiguous habitat patches across a landscape. In other words, local diversity may not mirror or represent regional diversity. There have been a number of suggestions that the creation of structural gaps through management systems can or should mimic the dynamics of natural gap creation through treefall, with the idea that community dynamics would thus be minimally affected. Other authors, however, have been less sanguine about the practicality of this.

Gaps and forest edges may often constitute habitats not unlike the forest canopy in respect of abiotic factors such as insolation, wind, temperature extremes, and rainfall, and may therefore exhibit an analogous flora and fauna.

D. Forest Fragmentation

The term *fragmentation* is often used to summarize the landscape-level structural changes to the forest exerted by a range of human activities. Forest fragmentation is a process that may take place over centuries, decades, or years. Often, this process starts with selective logging within a matrix of natural forest and may extend over time through the development of progressively more intensive agricultural landscapes (*agroscapes*). These agroscapes may eventually themselves become the matrix, finally leaving only isolated "islands" of residual forest. In developed countries, urbanization, suburbanization, "exurbanization," and the establishment of extensive transport infrastructure have all contributed to forest fragmentation, especially over the past 50 years. The process of forest fragmentation, and its effects on various aspects of biodiversity, has proven rich mine for ecological study. Several studies show that fragmented forests have less biodiversity than contiguous forests and that ecological and evolutionary processes that maintain biodiversity are compromised in forest fragments.

E. Spatial Mosaic of Forest Types

Both short- and long-term forest management actions have been shown to make radical changes in the spatial distribution of forest types and communities at the landscape scale. Natural forests are mosaics of species associations responsive in their distribution to aspect and degree of slope, altitude, exposure to winds and storms, fire history, and soil characteristics. A forest's spatial pattern and landscape context are thought to influence the regeneration of logged areas through seed and pollen dispersal dynamics and recruitment rates. Management can alter the natural forest mosaic by superimposing on it separate patterns reflecting land-use history, regeneration dynamics and history, or simply the pattern of access for logging (Cannon *et al.*, 1998; Foster, 1992). Altered distribution of forest types, even within a continuous forest cover, can have potentially harmful (or beneficial) effects on populations of some wildlife species.

F. Soil Effects

Effects of management systems on soils include alteration of the microstructure of large areas of forest soil through compaction and scraping, alteration of the forest floor profile through the creation of artificial pits and mounds (or the elimination of natural ones) by bulldozers, changes to the hydrology of the affected region, and losses of organic matter and nutrients that potentially threaten fertility of the site. The latter problem is perhaps most serious in dryland areas and some areas of the humid tropics, where soil structure is often less robust, lacking the highly developed humus layer and deep mineral soil profile of many forests in temperate climates. In some tropical and temperate rain forests, heavy rainfall tends to exacerbate soil problems through increased runoff from the exposed surfaces and compacted soils of roads and trails, the associated erosion of soil surfaces, and increased clogging and silting of lakes, streams, and wetlands. Old tractor tracks and landings may remain sources of direct runoff for decades, due to their low infiltration capacities. Amounts of sediment washing into streams have regularly been shown to be higher and more variable after logging disturbances. Increases in sediment off-flow by a factor of between 2 and 20 have been measured. These sediments often contain substantial amounts of organic material. Leaching of nutrients, as large volumes of rainwater fall directly onto exposed earth and percolate through the soil, is another problem, the severity of which depends both on the characteristics of the soil substrate and on the logging techniques employed. These soil changes may over time alter species composition and structure of forest vegetation at all levels.

The absence of downed logs and decomposing woody debris in temperate zone production forests has been blamed for reductions in soil fertility. A number of plant, animal, and fungi groups depend on complex forests floor structure for habitat and nutrient availability. All of these problems with soils have been said to be amenable to improvement through careful logging procedures. The U.S. state of New Hampshire and the Canadian province of British Columbia have each recently published suggested guidelines for maintenance of coarse woody debris in forestry operations.

IV. ABIOTIC CHANGES

A. Insolation, Temperature, and Wind Patterns

Direct structural changes may cause changes in a series of abiotic factors, again on several scales. Opening of the canopy alters the insolation regime in dense-canopied forests, though it may have less effect in forests with naturally sparse canopies. Sunlight reaching the forest floor has been measured at as little as 2% in closed-canopy tropical forests; selective logging raises this percentage manifold. Greater insolation has two kinds of effects: it increases the daily energy flux, especially to the lower strata of vegetation, and it raises the average temperature of both air and soil in the forest understory. Partial or complete removal of the forest canopy also increases the movement of the air column throughout the understory in response to winds above the canopy, affecting water retention, relative humidity, and transpiration rates in the understory. These changes affect the microclimates within the forest, making it warmer and drier on the average. This reduces the suitable habitat for some species of plants and animals, while increasing it for others.

B. Hydrology

Changes in forest floor hydrology may contribute to an increase in soil disturbances such as landslides and erosion, especially in the wet tropics, where rainfall is extreme. Such disturbances may increase the area of early successional vegetation; in areas where the organic component of the topsoil layer has been scoured away, the process of succession may be halted in its early stages for a long time. Even lesser levels of disturbance to the soil due to logging have been shown to reduce concentrations of CO_2, organic matter, and nutrients through leaching. Such changes will alter competitive balances among plant populations in favor of species adapted to disturbed habitats, and to the detriment of those specialized for the relatively consistent conditions of closed canopy forest. Some forest ecosystems are especially sensitive to changes in hydrology. Among these are mangrove forests, which depend on the regular flushing action of the tides and may be strangled by influxes of sediment from eroding soils. Lowland swamp forests, including Brazilian *varzea*, likewise may be gradually replaced by *terra firma* forests under conditions of sedimentation. Riparian forests of various kinds are important refuges for specialized plant communities, especially in arid regions. Scrub forest and other drylands forests are vulnerable to regional changes in rainfall patterns, since their plant communities may be at the edge of their ranges or environmental tolerances. Some of the most interesting and endangered forest types, in short, may be compromised by the hydrological changes that often accompany logging and other forest use.

V. BIOTIC CHANGES

Up till now we have been describing those measurable structural, abiotic effects of logging that have been documented in forests of various types worldwide. What effects do these changes have on the biodiversity, especially the species, which make up the forest?

Plants form the first, or *autotrophic*, layer of the forest ecosystem. The autotrophs are able to build organic compounds from inorganic building blocks, using energy from the sun. The other (*heterotrophic*) organisms in a forest are directly or indirectly dependent on the transformative power of photosynthetic plants for their livelihood. Therefore we will consider plants first. Fungi have been implicated in crucial symbiotic relationships with many forest plants and will be considered next. The most numerous *primary consumers* of plants in forests are the invertebrates. Primary consumers transform and recycle the organic compounds they derive from ingesting plants. This group will be discussed third. Some vertebrates, the group we review fourth, are also primary consumers (*herbivores*), others are secondary or tertiary consumers (*carnivores* and *omnivores*). *Detritivores*, which make up the final stage in the recycling of organic materials, are a very important component in forest ecosystems; we will discuss them in the context of effects on soils.

A. Plants

Since plants are stationary, they are most immediately influenced by changes in their abiotic surroundings. Such changes may influence plant species composition, density, evenness, community structure and associations, and plant interactions including mutualisms and parasitisms.

1. Early Successional Phases

A *pioneer guild* of plants is characterized by lack of shade tolerance and therefore tends to be favored in the light-rich microhabitats created by gaps. Pioneer species tend to be widespread, with airborne pollen and small airborne seeds, fast growth rates, and tolerance of a wide variety of environmental conditions, allowing them to disperse quickly into disturbed habitats. Hence, they are characteristic of managed forests, and the pioneer guild is relatively homogeneous in species composition over large regions (Finegan, 1996). This can be confusing to researchers, since localized increases of plant species diversity are common following logging. Logging expands the area of forest occupied by early successional phases, boosting representation of pioneer species in small samples. These local increases in diversity are not, however, matched on the regional level, since the local increases themselves come from the regional pool. Thus, logged forests tend to become increasingly homogeneous on a larger scale.

2. Exotic Invasive Species

Loss of tree cover often permits the recruitment of *naturalized exotics* (species that are not natural to the area), since the special conditions under which native species are competitively superior no longer exist. The presence of exotic plants tends to homogenize habitats and hence biodiversity. Examples of the many widespread and aggressive exotic plant problems are *Impatiens* (an Asian ornamental) in the neotropics, and the reed *Phragmites australis* in the United States.

B. Fungi and Mycorrhizae

A number of studies have shown that the presence in the soil of inoculate of certain species of symbiotic fungi is necessary for the regeneration and successful growth of a number of tree species, and the same may be true for other rooted plants. Soil mycorrhizae invade growing root tips and make available to them, in the course of their own metabolic processes, nitrogen and other nutrients that are otherwise locked up in the soil. Mycorrhizal fungi require certain soil conditions to flourish. If soil environmental conditions change radically, as can occur as a result of clear-cutting, for instance, these fungi may go locally extinct. Without healthy populations of the appropriate fungal symbionts, many tree species may be unable to regenerate successfully.

C. Invertebrates

We still know the least about the effects of logging on the first tier of forest-inhabiting animals, the invertebrates. One of the central unknowns is the number of invertebrate species inhabiting tropical forests, which is not known even to the nearest order of magnitude. Nevertheless, data on some taxa are available, and it is possible to distinguish some trends. Canopy denizens, especially flying insects such as butterflies and moths as well as flies, wasps, and bees, may be more easily seen or trapped in lower, disturbed forest, and hence may show up more easily in surveys. Like nectar-feeding and frugivorous bird species, some invertebrates may also benefit from the increased flowering and fruiting among early successional trees and vines in logged tropical forests. And many groups are adapted to the microclimates of forest edges and gaps, and so may benefit on the population level from structural changes and forest fragmentation. Nevertheless, lepidopterans and some dipterans and bee species are clearly reduced on a regional scale in disturbed tropical forests, because there are groups with very specific adaptations to the environmental conditions of closed canopy forest. For example, the hives of the large Asian honeybee (*Apis dorsata*) are usually found on very large old trees. Certain groups of tropical flying insects depend on vertebrate dung for part or all of their nutritional requirements. These may be strongly affected by reductions in mammal and bird populations that may accompany disturbance, although some may equally benefit from the presence of horses, cattle, or other domestic animals in the landscape. Other tropical winged insects are well adapted to the conditions of closed forest and have coevolved with particular resources that may be absent or rare in disturbed forest. Examples are euglossine bees and orchid flowers, parasitic wasps and fig fruits, or lepidopteran larvae. These groups may become rare when their hosts become rare.

In the temperate zone, lepidopterans, as well as other flying invertebrates, such as dragonflies, mayflies, and damselflies, are rarely found under closed forest canopy due both to their temperature requirements and to their dependence on bodies of water during part of their life cycle. Nevertheless, these too may be affected by uncontrolled logging activities that disturb their habitats within forests, such as bogs, swampy areas, ponds, and streams. Since boreal forests are relatively species poor, approaching the conditions of a monoculture, it is unclear to what extent cyclic events—such as insect infestations or fire—may be a part of their natural disturbance regime.

The invertebrate fauna of leaf litter and soil surface shows divergent responses to logging activities. Some groups (beetles, cockroaches, and millipedes) increase significantly in both species richness and abundance in logged tropical forests. Others (spiders, mites, scorpions, springtails, and termites) decrease. Still others, such as ants, seem to be reduced in species richness while remaining abundant in absolute numbers. Many detritivores (wood-boring beetles, certain termites, and soil mites as well as fungi) are habitat-dependent on standing or downed rotting wood. Such groups appear to have been negatively affected by long-term timber management in Europe and the American Southeast, where old growth forest structure including snags and large rotting logs has been effectively eliminated over large areas.

The effects of forest management on another important group of invertebrates, the parasites, are little understood. It is known that parasites are some of the strongest regulators of agricultural pests, and there are data showing that tropical agriculture, chronically vulnerable to pests, may actually be more successful where patches of natural forest are left between fields. Parasites and predators on insect pests may breed in forests and disperse into nearby fields, holding down pest populations.

It has often been observed that a dramatic increase in mosquito populations is a common phenomenon in logged areas, where drainage patterns are often affected. In the tropics this may constitute an important human health concern, and malaria is increasing in some areas.

In general, the invertebrates are still the least well known of the macroscopic fauna, yet their aggregate biomass is very large and their ecological influence is difficult to overestimate. E. O. Wilson has written about the essential role played by the "little things" in maintaining a variety of ecological functions (Wilson, 1987). The reduction or absence of groups of these forest denizens may have as yet unforeseen long-term consequences for forest ecology and ecosystem functioning.

D. Vertebrates

John Terborgh has observed that, in addition to the critical role played by Wilson's "little things" as the foundation of an ecological pyramid, there is an important top-down regulatory role played by the "big things," namely vertebrates and especially the large predators (Terborgh, 1988). In sites where top predators are absent, their prey populations, most of which are herbivores, tend to proliferate. This dynamic has been documented in tropical sites such as at the Smithsonian Institution Tropical Research Institute station on Barro Colorado Island, Panama, which lacks jaguar, in temperate parks devoid of wolf and puma populations, and in expanding suburbia all across North America, where white-tail deer and Canada geese are increasingly being viewed as pests by the human population. Many of the proliferating tropical herbivores, such as the large rodents agouti and capybara, are seed predators and may be affecting the regeneration of large-seeded tree species. In temperate parks, large deer and elk populations are browsing deciduous seedlings to the ground, effectively preventing their regeneration. The lack of top predators in managed forests may significantly alter forest structure and composition, both in the short and the long term.

Vertebrate populations in forests coming under management may be affected by losses or gains in (a) habitat area or (b) food resources. The direction of change is determined both by the particular needs of each species and by the nature of the management system. The many permutations of these variables explain why we see divergent responses to management among vertebrates.

1. Habitat Loss or Increase

The loss of important habitat elements tends to affect the more specialized and the less mobile species most strongly, since they may not be able to disperse to new habitats. Loss of the appropriate microclimate, sufficient kinds and amounts of cover, nesting sites, or even perch sites may force species to local extinction. We know that some vertebrates depend on habitat elements found only in old, undisturbed forests. A well-known example is the northern spotted owl in the Pacific Northwest of the United States, which nests in the large hollow trees that are usually eliminated by timber management. A range of large temperate and tropical birds, including large woodpeckers, hornbills, and quetzals, likewise nest exclusively in hollows and may use and reuse the same holes year after year. Some mammals, too, are critically dependent on old-growth habitat elements for survival. The American marten, for instance, which has very little body fat yet remains active throughout the winter in cold and snowy environments, needs resting places below the surface of the snow, often under rotting stumps or large logs. Thus, although animals like the omnivorous marten may benefit from the increased seasonal availability of berries or other resources in logged forest gaps, their territories must also include areas of older forests with woody debris.

It has been shown that a number of smaller forest birds, both temperate and tropical residents, avoid gaps in forest cover such as those created by logging. Some are reluctant to cross gaps of as little as a few tens of meters. This may result from fear of predation or from physiological adaptation to a narrow range of understory temperatures and humidity levels. Under the increasingly predominant conditions of forest fragmentation, these species may find it difficult to locate appropriate habitat, even when it is available nearby.

Terrestrial amphibians, including frogs and salamanders, may often find the relatively desiccating conditions of open-canopy disturbed forest too warm and dry. Because they "breathe" partly by the exchange of gases directly across the skin, amphibians are particularly vulnerable to alterations in their microclimates. One element in the present worldwide crisis among amphibian populations may be reductions in suitable cool, moist, and shaded old-growth habitat.

On the other hand, some vertebrate populations may do well in disturbed habitats in secondary succession.

Examples of this group that have been studied are mice in temperate forest fragments, tenrecs in Madagascar, and other rodents in tropical and temperate forests, all of which may benefit from the increase in undergrowth and herbaceous ground cover accompanying management.

2. Loss or Increase of Food Resources

Opening the forest canopy may often stimulate a flush of new leaves or fruits as the trees and understory respond to the more generous light allowance. Many browsing and fruit-eating vertebrates benefit from this change and may seek out areas of secondary succession such as logged areas. Populations of large folivorous mammals, including elephant, tapir, duiker, gaur, bearded pig and peccary in the tropics, or deer and elk in the temperate zone, often do well in secondary forest. In the absence of predation, including human hunting, they may themselves become agents of serious ecological change by grazing and trampling large areas of forest understory. On the other hand, many of these large browsing mammals are important game animals, especially in tropical forests where hunting is particularly difficult to control. Many have evolved shyness responses to humans, and most have greatly reduced populations in areas where hunters have access.

3. Birds

The effects of logging on the species richness of birds are highly variable. In logged tropical forest, certain feeding guilds experience a reduction in population ranging from 25% up to 100%. Hard-hit guilds include terrestrial foragers (insectivores and frugivores) and the small insectivorous species associated with understory habitats, particularly those of the forest interior. These decreases are likely the result of reductions in the insect resource base or microhabitat changes. These findings are not completely consistent among studies, however. Some studies, for instance, have found small increases in abundance for terrestrial, understory, and foliage-gleaning species after logging in Malayasia. Substantial local increases in abundance have also been noted for generalist feeders that can supplement nectar and fruit resources with insects. The opening of the canopy often, though not always, increases the availability of nectar and fruits in early successional patches, in turn supporting more abundant populations of generalist feeders.

4. Primates

Primates have particularly complex responses to logging and to the presence of humans in managed forests. In the Neotropics, generalist feeders (*Pithecia* spp., *Saguinus* spp.) have been found to do well in disturbed areas, whereas frugivores (e.g., *Chiropotes* spp., *Ateles* spp.) tend to leave logged-over areas, especially when their food trees have been extracted. However, there are exceptions, such as *Callicebus torquatus*, a generalist feeder that was found to be absent from logged forest. Old World frugivores such as *Pongo pygmaeus* and *Pan troglodytes* decline sharply in population density after logging, whereas folivores (e.g., *Colobus* spp., *Gorilla gorilla*) may seek out disturbed forest for the leaf flush.

E. Behavioral Changes

Behavioral changes among vertebrate populations have received little attention until recently. There are substantial difficulties inherent in studying and quantifying them. Nevertheless, they should be considered a significant group of changes because their effects may be delayed until one or several generations after the disturbance itself. Examples, which have been noted, are changes in social organization and breeding behavior among primates in Malaysia, and in nesting behavior and dispersal among birds in fragmented habitats. Among animals with complex response patterns, there are doubtless many other instances yet to be studied of behavioral changes stemming from human interventions. These may affect population dynamics years or decades after an event such as logging.

F. Domestic Animals and Nonforest Species

Domestic animals and introduced species have been important influences on forests throughout history. In European oak and beech forests, for instance, pigs and other grazing animals were grazed in commons forests on a seasonal basis to take advantage of fruit masting. Sheep and goats have virtually eliminated forest regeneration on the once lushly tree-clad slopes of the Mediterranean and throughout the Middle East (Thirgood, 1981). Today, cattle browsing is a major influence on remaining forest fragments in many areas of the tropics.

The movement of nonforest species into logged and disturbed forests is the origin of some of the apparent increases in species diversity in logged forest plots. More species of coleoptera, for instance, have been counted in logged tropical forests than in neighboring primary forest. The species lists in logged areas, however, included the names of field species that had entered the forest after the canopy had been degraded.

G. Aquatic Fauna

The loss of aquatic biodiversity is understudied in the tropics. Concerns regarding changes in hydrology resulting from forest disturbance overlap with local community concerns, since the need to protect sources of clean water is becoming acute in many places as human populations grow. There is more information about the effects of logging on streams and aquatic life in the temperate zone. The removal or degradation of tree cover on slopes, including the construction of logging roads, often results in topsoil loss and erosion. Soil is washed downhill and ends up in streams and rivers. Silt deposition causes declines among anadromous fishes, bottom-feeding river fishes, bivalves, and other invertebrates. Heavy silt loads can have major impacts far downstream—even in some cases degrading coastal habitats such as tropical coral reefs, mangroves, and offshore fisheries. Changes in the temperature of stream water as a result of streamside vegetation loss can also have far-reaching effects on faunal diversity.

VI. ECOLOGICAL INTERACTIONS

Structural, abiotic, and biotic changes are likely to alter a range of ecological interactions and evolutionary processes. Changes in hydrology, soils, and soil fauna, for example, should influence nutrient dynamics, which may in turn have further impacts on biodiversity. Biotic changes alone can alter species interactions. Again little is known about the impact of extractive regimes on ecosystem function and ecological interactions. The impacts are likely to be most dramatic where extracted species constitute dominant elements of forest ecosystems as, for example, the dipterocarp tree species in Southeast Asian dipterocarp forests. Indeed, much of the recent evidence that logging may influence vital interactions and ecological processes comes from such forests.

Pollination, seed dispersal, and seed predation are basic processes that profoundly influence reproductive output and regeneration of plants. By lowering the density of harvested species, logging can decrease the resources available to pollinators, animal seed dispersers, and seed predators. The vast majority of tree species are outcrossed, and in tropical forests, where most tree species are pollinated by animals, substantial declines in the abundance of food plants can result in changes in pollinators' foraging behavior. One study in Southeast Asian *dipterocarp* forests confirmed that logging in forests containing *Shorea siamensis* increased the average distances among the remaining trees of that species. The large distances made it difficult for the bees that pollinate *S. siamensis* to move from tree to tree, concomitantly limiting outcrossing potential. Opening up the canopy also led to colonization of the forest floor by flowering plants that attracted bees to the understory, further depriving the canopy trees of their pollinators (Ghazoul, 1999).

Reduction in the density of reproductive individuals can also result in a substantial decline in the number of seedlings. A recent study compared seedling production in logged and unlogged dipterocarp forests of Indonesia (Curran *et al.*, 1999). Logging reduced the number of reproductive individuals per hectare to 3% of the original. Seedling production in the logged forest, following a mast fruiting year, was a mere 15% of that in the nearby unlogged forests.

The same study also documented indirect effects of logging. Dipterocarps undergo mast fruiting every few years associated with local El Nino-Southern Oscillations (ENSO). Logging around the Gunung Palung National Park, the site of the study, has fragmented the once contiguous forest. Changes in land use have apparently affected local ENSO climate conditions. As a result, spatial synchrony in mast fruiting and the level of seed production have been reduced. One result of the natural mast fruiting pattern is that copious amounts of seed are produced simultaneously. It is assumed that this "satiates" seed predators, leaving a substantial surplus of intact seeds for regeneration. However, with the reduction in seed production, seed predators are not getting "satiated," and there is little seed surplus. Even in the national park, where there is no logging, there has been inadequate regeneration because of the decrease in seed production and the movement of seed predators into the park from surrounding areas lacking mature reproductive trees.

These two studies from Southeast Asian dipterocarp forests illustrate several of the ways in which logging can have far-reaching and perhaps unexpected effects on forest dynamics. (a) Removal of mature dipterocarps and the opening of the canopy (both structural changes) produce new microclimatic conditions and an altered array of understory species (a biotic change). (b) These changes combine to produce altered foraging patterns among bees, compromising their efficiency as pollinators—an essential ecological interaction. (c) Meanwhile, fragmentation and degradation of the forest cover alters local climatic patterns (an abiotic change), which again alter an ecological process, the mast-fruiting pattern. (d) This alters seed predation dynamics and affects regeneration, potentially influencing forest structure and composition into the long term.

Considerably more work is needed to fully understand the consequences of the altered ecological relationships brought about by logging and other extractive activities. Studies on the effects of logging have largely been restricted to the impacts on structural and biotic components. Removal of nontimber forest products may also be expected to have an impact on a diverse range of ecological interactions and processes but has received even less systematic study.

VII. GENETIC EFFECTS AND EVOLUTIONARY PROCESSES

Alterations in density of tree species may also have a profound effect on evolutionary processes. Most trees are strongly outcrossed; outcrossing rates are highly dependent on population density. Low population density can restrict the movement of pollen among trees and result in a high level of inbreeding. Although it has been difficult to document the deleterious consequences of a decrease in outcrossing rate, it has been confirmed in several species that inbreeding does increase with a decrease in density.

In the tropics where many tree species typically occur in low densities, reduction in density due to extraction may further decrease population sizes. Small population sizes may also contribute to inbreeding. Thus, inbreeding may be increased both by the effects of decreases in density on the mating system and by decrease in overall population size.

Extraction can also remove the better-adapted genotypes from the population, leading to *dysgenic* selection, or selection for less well-adapted genotypes. Populations subject to harvest have lower levels of heterozygosity and lower overall levels of genetic variation. Inbreeding in logged populations resulting from a decrease in density may further decrease overall levels of heterozygosity.

Genetic and evolutionary effects can be cumulative. Fragmentation, decrease in density, and reduction in effective population size all contribute to inbreeding. In trees that are highly outcrossed, increasing levels of inbreeding may be particularly deleterious in the face of a changing environment. The genetic consequences of management interventions in forest ecosystems remain poorly explored.

VIII. SYNERGISTIC ANTHROPOGENIC EFFECTS

Probably only a minority of extractive forest operations today are performed under a management plan, and the vast majority of these are in the developed, less biodiversity-rich countries. In the tropics, in contrast, only about 2% of forests are currently under any active management plan, according to one authoritative study (Poore *et al.*, 1989), although most are experiencing extractive pressure. The absence of planning and long-term control most often results in a series of secondary effects, including increased hunting, fire, invasion by exotic species, mining activities, shifting agriculture, and illegal logging, which may far outweigh the effects on biodiversity of the original management operations.

It has been noted that large amounts of secondary vegetation and brush tend to alter the understory structure of cut-over areas. The role of such vegetation and brush in fueling forest fires, both in temperate forests and even in the relatively fire-resistant moist tropical forests, has received increasing attention in recent years (e.g., Nepstad *et al.*, 1999). It is clear that large-scale fires are altering microclimatic conditions over wide areas of tropical Asia and South America. In some of the drier temperate forests, for instance parts of the U.S. West and the pine barrens of the Atlantic seaboard, vegetation communities have evolved under the influence of regular low-intensity fires. The suppression of such fires during this century has altered the makeup of these communities and in some cases lowered their resistance to large destructive conflagrations. Reintroduction of controlled fire regimes is beginning to allow the native species to return and compete successfully with exotic species. A somewhat contrasting case is found in certain areas of the dry tropics, which are, as a group, some of the most highly altered and degraded forest ecosystem types. Here, degradation of the native plant assemblages and hydrology have virtually eliminated the forest's ability to regenerate itself in the face of continual incursions of fire. In such areas, putting a stop to forest fires over at least the next several decades is seen as a key step to restoring the natural plant communities.

Sharp increases in hunting after logging, too, have become the object of attention recently (Oates 1999; Robinson *et al.*, 1999; Struhsaker, 1997). Logging requires roads. Particularly in the tropics, where hunting is often unregulated, logging roads provide easy access for hunters and their vehicles into previously remote areas. Logging crews themselves are often the first hunters in a newly accessed area, both for their own consumption and for commercial markets. Others quickly follow them, and pressure on the game animal species can become intense.

In many parts of the tropics, one of the most intractable effects of logging is settlement by colonists eager for new agricultural land. Colonists follow the logging crews, entering the forest along logging roads, carving out homesteads, and planting crops. Since such movements tend to fragment and finally convert forest to agricultural land, they are clearly some of the most deleterious influences on biodiversity. Further extractive activities associated with wresting a living from the forest, such as mining, fishing, and trapping animals and birds, tend to accompany colonization and increase pressure on forest resources in a multitude of ways.

IX. CONCLUDING REMARKS

Clearly the extraction of forest products in managed ecosystems has an impact on biodiversity. However, natural ecosystems are dynamic and subject to all sorts of perturbations. The critical questions in the case of managed ecosystems are the extent to which biodiversity is lost and the degree to which the losses are irreversible. A landscape perspective is important for addressing these questions, because biodiversity may increase locally while decreasing at the regional level. Moreover, some local changes may be a manifestation of regional level changes and vice versa. The landscape perspective is also important to unraveling the confounding but synergistic effects of habitat fragmentation and climate change on biodiversity, as is evident from the recent work of Curran and colleagues in Southeast Asia (see also Bawa and Dayanadan, 1998). A wider outlook is also critical to understanding the impacts on the ecological and evolutionary processes and the ecosystem services that have been so far ignored in studies concerned with biodiversity in managed ecosystems. Finally, we do not imply that biodiversity in managed forests will always substantially decline. The low-impact harvesting regimes that are being explored all over the world may allow significant and critical goods and services to be conserved and sustainably used, particularly if such management systems are well integrated with their social and culture milieus and are responsive to the social needs of specific regions.

Acknowledgments

This represents publication number 75 of a program in biodiversity conservation supported in part by the MacArthur Foundation, involving the Ashoka Trust for Research in Ecology and the Environment (ATREE), Bangalore; the University of Agricultural Science, Bangalore; Pondicherry University, Pondicherry; the Institute of Social and Economic Change, Bangalore; and the University Massachusetts at Boston.

See Also the Following Articles

AGRICULTURE, SUSTAINABLE • DEFORESTATION AND LAND CLEARING • FIRES, ECOLOGICAL EFFECTS OF • FOREST CANOPIES, ANIMAL DIVERSITY • FOREST ECOLOGY • PREDATORS, ECOLOGICAL ROLE OF • TIMBER INDUSTRY

Bibliography

Bawa, K. S., and Seidler, R. (1998). Natural forest management and conservation of biodiversity in tropical forests. *Conservation Biology* 12, 46–55.

Bawa, K. S., and Dayanandan, S. (1998). Climate change and tropical forest genetic resources. *Climate Change* 23, 449–466.

Bouman, O. T., and Brand, D. G. (Eds.) (1997). *Sustainable Forests: Global Challenges and Local Solutions.* Food Products Press, New York.

Bruenig, E. F. (1996). *Conservation and management of tropical rainforests: An integrated approach to sustainability.* CAB International, Wallingford.

Cannon, C. H., Peart, D. R., and Leighton, M. (1998). Tree species diversity in commercially logged Bornean rainforest. *Science* 281, 1366–1368.

Curran, L. M., Caniago, I., Paoli, G. D., Astianti, D., Kusneti, M., Leighton, M., Nirarita, C. E., and Haeruman, H. (1999). Impact of El Nino and logging on canopy tree recruitment in Borneo. *Science* 286, 2184–2188.

Finegan, B. (1996). Pattern and process in neotropical secondary rain forests: The first 100 years of succession. *Trends Ecol. Evol.* 11, 119–124.

Forman, R. T. H. (1995). *Land Mosaics: The Ecology of Landscapes and Regions.* Cambridge University Press, Cambridge.

Foster, D. R. (1992). Land-use history (1730–1990) and vegetation dynamics in central New England, U.S.A. *Journal of Ecology* 80, 753–772.

Frumhoff, P. (1995). Conserving wildlife in tropical forests managed for timber. *Bioscience* 45, 456–464.

Ghazoul, J. (1999). Direct and indirect effects of human disturbance on the reproductive ecology of tropical forest trees. *Plant Ecology* (in press).

Grieser Johns, A. (1997). *Timber Production and Biodiversity Conservation in Tropical Rain Forests.* Cambridge University Press, Cambridge.

Hunter, M. L. (Ed.) (1999). *Maintaining Biodiversity in Forest Ecosystems.* Cambridge University Press, Cambridge.

Kirby, K. J., and Watkins, C. (Eds.) (1998). *The Ecological History of European Forests.* CAB International, Wallingford.

Laurance, W. F., and Bierregaard, R. O. (1997). *Tropical Forest Remnants.* University of Chicago Press, Chicago.

Nepstad, D. C., Verissimo, A., Alencar, A., Nobre, C., Lima, E., Lefebvre, P., Schlesinger, P., Potter, C., Moutinho, P., Mendoza, E., Cochrane, M., and Brooks, V. (1999). Large-scale impoverishment of Amazonian forests by logging and fire. *Nature* 398, 505–508.

Oates, J. F. (1999). *Myth and Reality in the Rain Forest: How Conservation Strategies are Failing in West Africa.* University of California Press, Berkeley.

Poore, D., Burgess, P., Palmer, J., Rietbergen, S., and Synnot, T. (1989). *No Timber Without Trees: A study for ITTO.* Earthscan Publications, London.

Richards, P. W. (1996). *The tropical rainforest: an ecological study*, 2nd ed. Cambridge University Press, Cambridge.

Robinson, J. G., Redford, K. H., and Bennett, E. L. (1999). Wildlife harvest in logged tropical forests. *Science* 284, 595–596.

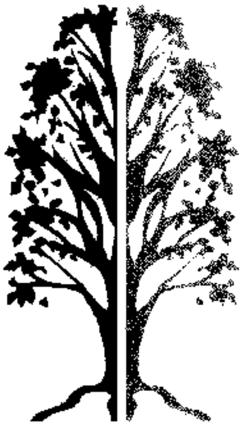
Schelhas, J., and Greenberg, R. (1996). *Forest Patches in Tropical Landscapes.* Island Press, Washington, D.C.

Struhsaker, T. T. (1997). *Ecology of an African Rain Forest: Logging in Kibale and the Conflict between Conservation and Exploitation.* University Press of Florida, Gainesville.

Terborgh, J. (1988). The big things that run the world—A sequel to E. O. Wilson. *Conservation Biology* 2, 402–403.

Thirgood, J. V. (1981). *Man and the Mediterranean Forest: A History of Resource Depletion.* Academic Press, London.

Wilson, E. O. (1987). The little things that run the world. *Conservation Biology* 1(4), 344–346.



LOSS OF BIODIVERSITY, OVERVIEW

Robert Barbault
Institut Fédératif d'Écologie Fondamentale et Appliquée
Université Pierre et Marie Curie

GLOSSARY

biodiversity The variety of organisms considered at all levels, from genetic variants within the same species to the whole range of species and ecosystems.

extinction The disappearance of any lineage of organisms, from populations to species and higher taxonomic categories (genera, families, phyla). It can be local or global (total).

genetic drift The process of random sampling of genes that leads to changes in the genetic composition of a population. The effect of this process is particularly important at small population sizes.

homozygous Possessing the same gene form (allele) on both chromosomes.

locus The position on a chromosome of a gene. It is determined by any number of allelic forms.

metapopulation The set of populations (or subpopulations) of the same species, linked by migration. It is characterized by the processes of local extinction and recolonization.

Pleistocene The geological time that ends with the last glacial period and the appearance of humans. It started 2 million years ago and finished 10,000 years ago with the end of the last Ice Age.

stochastic The result of chance and random processes.

THE EXTINCTION OF A SPECIES, like the death of an individual, is a natural phenomenon—its inevitable destiny. In fact, during the long history of life, earth has experienced several periods of mass extinction. But the crisis of current extinction differs from the preceding ones in that it is the direct result of human activities. By our ecological success—haven't we invaded the entire planet?—amplified by our industrial and technological revolutions, our species exerts such an impact on the biosphere that one witnesses today an acceleration of extinction phenomena without precedent. This erosion of biodiversity is expressed on three interdependent levels; it affects (a) the diversity of our planet's ecosystems and landscapes, (b) the richness of species in the faunas and floras in most parts of the world, and (c) the genetic diversity of many natural and domesticated species. The objective of this chapter is to address the patterns, causes, and extent of such losses of biodiversity.

I. LESSONS FROM THE PAST

The longest phase in the evolution of life on our planet extends for the 2 billion years from the appearance of the simplest molecules capable of autoreplication to the appearance of the first prokaryotes. Unfortunately, nothing is known about the evolution of biodiversity throughout this very long period, dominated by microorganisms. Our attention concerning extinction must therefore be restricted to two kingdoms, the plants and the animals or, in other words, to the past 600 million years.

A. Extinction Rates Deduced from the Fossil Record

Table I shows the estimates for the average life span of species in various groups of fossils from their origin to their disappearance. Overall, the average life span of the species is about 5 to 10 million years.

Based on this estimate, and remembering that the total interval of time considered is equal to 600 million years, one would estimate that the present stock of animals and plants accounts for only approximately 1 to 2% of all species that have ever lived. However, this evaluation obviously depends on the quality of the fossil record. One should also note that the estimated life spans vary considerably according to the groups considered (Table I). In particular, according to Raup, the average duration of mammal species is on the order of 1 or 2 million years, much lower than the 10 million years estimated for (mainly marine) invertebrates. Taking account of this variability, Sepkoski (1992) estimated that the current species of plants and animals account for approximately 2 to 4% of those that have ever existed.

TABLE I

Estimates of Species' Life Spans in Various Groups from Origination to Extinction[a]

Group	Species' average life span in millions of years
All invertebrates	11
Marine invertebrates	5–10
Marine animals	4–5
Mammals	1–2
Diatoms	8
Dinoflagellates	13
Planktonic Foraminifera	7
Bivalves	10
Echinoderms	6

[a] Based on data from various sources (from R. May *et al.*, in Lawton and May, 1995, simplified).

Rates of extinction tend to vary considerably over time (Fig. 1). Thus, four of the five major extinction periods in the fossil record (Fig. 1) each eliminated between 65 and 85% of the marine animal species (Ordovician, Devonian, Trias, and Cretaceous), and one (Permian) eliminated 95% or more.

B. Mass Extinctions Since the End of the Pleistocene

During the last 50,000 years, several extinction episodes have concerned almost exclusively the terrestrial mammals, some birds, and, with a lesser degree, some reptiles. A comparison of the data gathered on the various continents makes it possible to refine the explanatory assumptions (Fig. 2). Africa gained and lost several genera of mammals throughout the Pleistocene up to the Holocene, 10,000 years ago, but was left with a primarily intact fauna. In Europe, the extinctions touched only the largest mammals (>44 kg). Australia was subjected to a massive peak of extinctions, with 86% of the large-bodied animal genera, including the large reptiles and giant birds, going extinct at the end of Pleistocene. The majority of extinctions of the megafauna occurred about 30,000 years ago. Finally, in North America mass extinction occurred between 10,000 and 12,000 years ago. Of the 37 kinds of mammals that died out, 33 weighed more than 44 kg. Similarly, extinctions in South America affected only the genera of animals exceeding 44 kg.

Two classes of interpretations have been advanced to explain these extinctions, one focusing on human activities as an extinction force, the other focusing on climate changes. Thus, the episode of extinction of North America occurred at the end of Pleistocene, when temperatures rose, leading to changes in the composition and the structure of plant communities, thus disturbing the interactions between plant and their browsers, and consequently leading to the extinction of many herbivores (particularly the large ones) as well as the predators that depended on them. Similar arguments may be used in Australia, where the climatic change at the end of the Pleistocene was associated with a higher frequency of drought periods. However, the asynchrony of the extinction crises in America and Australia may lead one to favor the responsibility of humans, since the colonization of these continents by humans was also asynchronous.

The well-documented case of Madagascar is particularly interesting. Humans arrived there 1500 to 2000 years ago. At the beginning of the Christian era, 17 species of lemurids inhabited the island. Seven of them disappeared about 1000 years ago; all but one were

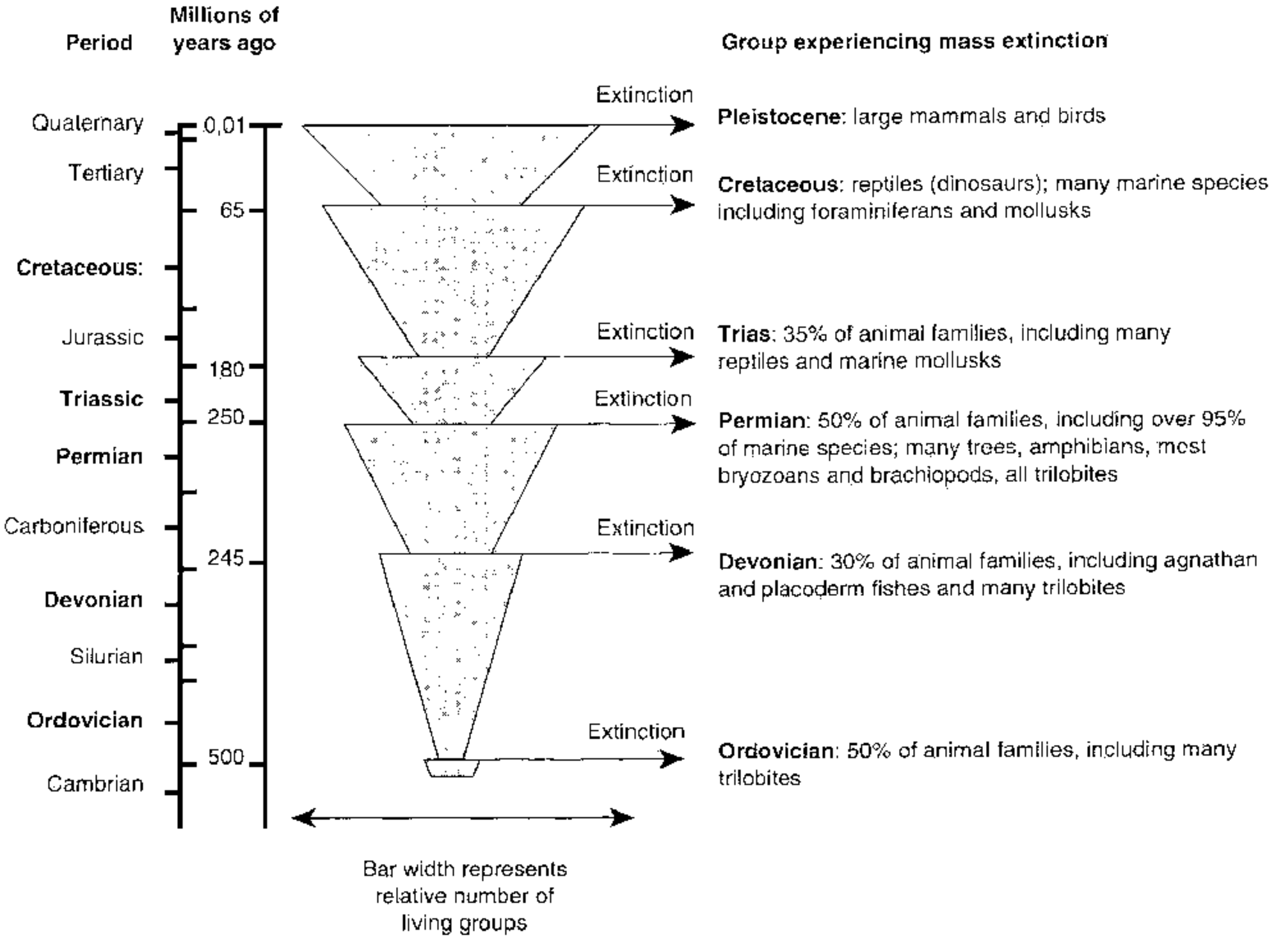


FIGURE 1 Although the total number of families and species has increased over the eons, during each of five episodes of natural mass extinction a large percentage of these groups disappeared. We are now at the start of a sixth episode, the Pleistocene extinction, as human populations eliminate species through habitat loss and overharvesting. From Primack (1993).

large species, probably diurnal and mostly terrestrial. The surviving species were, in contrast, either nocturnal or small, arboreal species. Three species of mammals from other families disappeared at about the same time: a pigmy hippopotamus, *Hippopotamus lemeriei*, and two large carnivores, *Cryptoprocta spelea* and *Plesiorycteropus madagascariensis*; furthermore, two species of the giant tortoises, *Geochelone* and the elephant bird *Aepyornis* (which is, with a weight of 500 kg, the largest known bird), went extinct. While it is not clear whether these extinctions are caused by the direct effects of hunting or by a degradation of the environment, it is certain, however, that they are linked to human activities.

II. CURRENT EXTINCTION RATES

With the increasing concerns about the erosion of the biodiversity and the developing polemic about the Convention on Biological Diversity since the Earth Summit of Rio in June 1992, it is becoming more necessary to have scientific information about the current rates of extinction. From the development carried out on the initiative of John Lawton and Robert May (1995), it is possible to draw some general conclusions.

A. Recent and Current Extinction Rates

The available data, summarized in Table II, show a number of points of general interest. Half of the extinctions listed since 1600 occurred during the 20th century; the majority affect species inhabiting islands. Extinctions among the insects are much rarer, relative to the number of known species, than among vertebrates (0.006 versus 0.5%). Although we have noted that the average life span of insect species, estimated from the fossil records, could be 10 times higher than that of

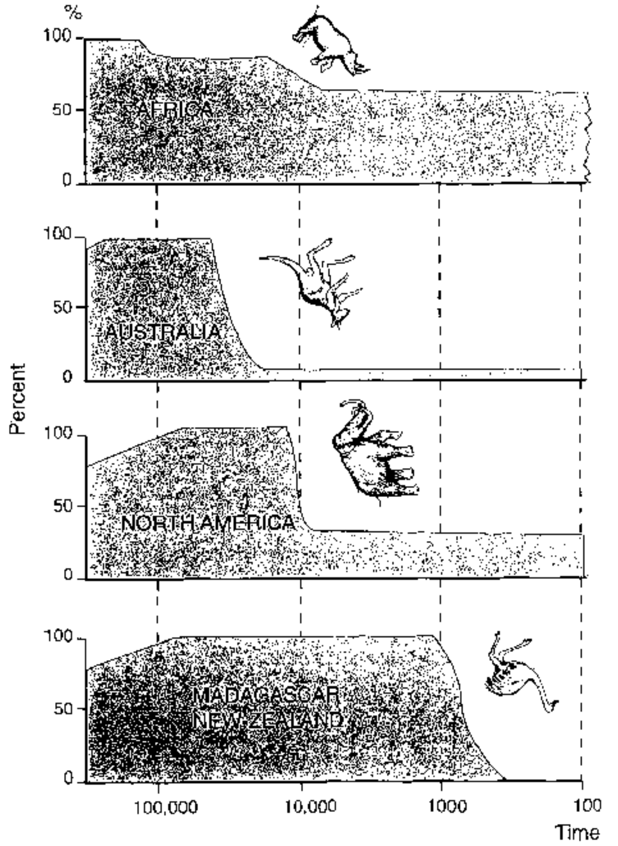


FIGURE 2 The extinction of large vertebrates coincided with the arrival of humans in Australia, North America, Madagascar, and New Zealand. In Africa, where humans and animals evolved together for millions of years, the damage was less severe. From Wilson (1992).

tetrapods, the difference in extinction rate of two orders of magnitude is difficult to explain.

Of course, the differences in extinction rates among groups could result partly from the difference in interest granted to one or the other group. Thus, the number of taxonomists working on vertebrates is 100 times the number working on invertebrates, and 10 times higher for vertebrates than for the vascular plants. This differential attention appears in the rates to which new species are discovered: three to five new species of birds are discovered each year (i.e., 0.03 to 0.05% of the known total), while the tropical botanists can hope to find a new species of plant among every 100 specimens; collections of insects, mushrooms, or marine microfaunas carried out in unstudied areas reveal proportions of new species ranging from 20% to 80%.

A final point worth emphasizing is that, even at groups known comparatively well such as the birds and the mammals, the listed extinctions are certainly underestimated.

B. Estimate of the Future Rates of Extinction from the Area—Species Richness Relationship

Future rates of extinction can be estimated by merging the estimates on the losses of the environment (deforestation, for example) with the expected relationship between area and species richness. This relationship is an empirical rule, based on a number of studies, that connects the number of species S for a given taxonomic group (coleopterans, birds, vascular plants, etc.) and the area A of the "islands" (true islands, or ecologically isolated areas such as lakes, mountain peaks, clearings in a forest) inhabited by the species (Fig. 3). Generally, the relationship between species richness and the area has the form:

$$\log(S) = c \log(A)^z$$

where c is a constant and the parameter z is usually between 0.15 and 0.35. Assuming that this equation can also describe the reduction of species richness driven by deforestation or other processes decreasing the area of suitable habitat, and taking into account the current rate of deforestation in the tropics (which varies between 0.8 and 2% per year), one can deduce an annual rate of extinction of between 0.2 and 0.5%. In other words, with a total richness of currently 5 million species inhabiting the planet (our minimal estimate), between 10,000 and 25,000 species will disappear every year.

However, this kind of extrapolation is associated with several problems. It is not known, for example,

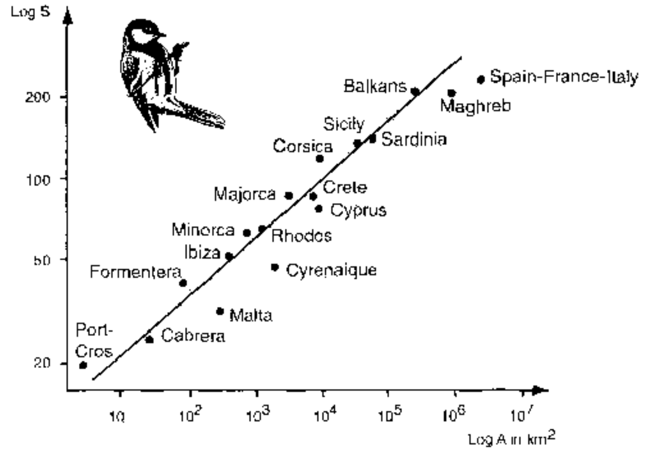


FIGURE 3 An example of the log-log relationship between the number of breeding-bird species (S) and the area of islands and continental regions in the Mediterranean (A). From Blondel (1995).

TABLE II

Species in Major Taxa That Have Become Extinct since 1600 or Are Now Threatened with Extinction[a]

Group	Number of extinctions	Number of species threatened[b]	Approximate recorded species (in thousands)	Percentage extinct	Percentage threatened
Animals					
Molluscs	191	354	100	0.2	0.4
Crustaceans	4	126	40	0.01	3
Insects	61	873	1000	0.006	0.07
Vertebrates	229	2212	47	0.5	5
Fishes	29	452	24	0.1	2
Amphibians	2	59	3	0.1	2
Reptiles	23	167	6	0.4	3
Birds	116	1029	9.5	1.2	11
Mammals	59	505	4.5	1.3	11
Plants					
Gymnosperms	2	242	0.8	0.3	32
Dicotyledons	462	17,474	190	0.06	9
Monocotyledons	120	4421	52	0.9	9

[a] Reprinted with permission from *Nature* (Smith *et al.*, 1993) Copyright 1999 Macmillan Magazines Ltd.

[b] "Threatened" includes the IUCN (World Conservation Union) categories of "vulnerable," "endangered," and "probably extinct."

up to what point the effects of the fragmentation of the environment, particularly in the tropics, can be deduced from the area-species relationship that has been described mainly for insular biogeography. Thus, Simberloff stressed that, during the past 2 centuries, only three species of birds are known to have gone extinct in the forests in the eastern United States, even though the forest has been reduced to small fragments with a total of only 1 to 2% of their original area.

In spite of such uncertainties about our estimates, it is certain that the reduction and the fragmentation of the environment, even if not always resulting in a reduced density of populations, will increase the risk of their becoming extinct (Table III in Box 1).

III. MECHANISMS OF EXTINCTION

What aspects of the environmental conditions make a species or population go extinct? It is clear that, if the mortality rate continuously exceeds the birth rate, a species will eventually become extinct. However, any

TABLE III

Number of Species Considered to be "Threatened" by the World Conservation Monitoring Centre (the WCMC consideres all species classified as "endangered," "vulnerable," "rare," or "indetermined" as threatened).

Group	Endangered	Vulnerable	Rare	Indetermined	Total
Mammals	177	199	89	68	433
Birds	188	241	257	176	862
Reptiles	47	88	79	43	257
Amphibians	32	32	55	14	133
Fishes	158	226	246	304	934
Invertebrates	582	702	422	941	2647
Plants	3632	5687	14.485	5302	26,107

Box 1

Future Extinction Rates Deduced from IUCN Red Lists

The International Union for the Conservation of Nature (IUCN) proposed a number of criteria to classify the level of threats imposed on plant and animal species. Seven categories are defined:

Extinct: A taxon that has not been observed in nature for more than 50 years.

Endangered: A taxon close to extinction and whose survival is at risk if no action is taken to eliminate the causes of its disappearance. Included in this category are species whose populations are reduced to a critical level or whose habitats are threatened.

Vulnerable: A taxon at risk of entering the "endangered" class if detrimental factors continue to exert their effects. This category includes those taxons whose populations decrease as a consequence of overexploitation, large-scale habitat destruction, or any other perturbation of their environment, and those taxons whose populations remain abundant but that are nonetheless threatened by a variety of detrimental factors.

Rare: Taxa whose populations are globally rare even though they are not endangered nor vulnerable. Such taxons are generally found in very specific areas or habitats or have widely dispersed small populations.

Indetermined: A taxon that has to be included in one of the above categories, but about which insufficient information is available to assess the level of risk.

Insufficiently known: A taxon thought to belong to one of the preceding categories, but without data to substantiate this.

Threatened: A taxon included in any of the preceding categories.

On this basis it appears that the number of animal species cited as threatened has grown by more than 30% between 1986 and 1990. During this interval, 15 vertebrate species (33 animal species in all) have been added to the list of recent extinctions; if it continues unabated, this would result in the extinction of half of the 47,000 vertebrate species in 700 years. Similarly, between 1990 and 1992, 163 plant species have been added to the list of extinct species; this corresponds with the extinction of half of the earth's 250,000 plant species in approximately 3000 years.

Though all of these estimates remain fragile, the table of threatened species (Table III) constitutes a useful starting point for conservation strategies.

prediction (which is based on perhaps a decade of observation) about extinction during the next, say, 3000 or 7000 years assumes that nothing will change. Since we are assuming that the major cause of extinction is humankind, this assumption appears unlikely.

Therefore, we will discuss more short-term risks of extinction, risks associated with an environment where, in principle, the population should persist (where births exceed deaths, at least at low population densities; see Box 2). In doing so, we will distinguish isolated populations (a situation which is rapidly developing with the human transformation of landscapes) and interconnected populations constituting so-called metapopulations. The main difference between the two is that (unless due to human intervention) only interconnected populations can benefit from immigration.

A. Isolated Populations

1. Low Population Densities: The Allee Effect

At low densities, individuals may be distributed over distances that are much larger than the distances they usually move in, so that the probability of meeting one another for mating is small. Therefore, the growth rate of the population may drop below unity when the density is below a certain threshold. A well-established example of such an Allee effect leading to the extinction is the mottled woodpecker *Dendrocopos medius* in Sweden in 1982. It is probable that an Allee effect is also responsible for the extinction of the migrating pigeon of North America, formerly so spectacularly abundant.

The possibility of extinction due to low population densities is particularly alarming for large animals, which live naturally at a low density, and for species with marked sociality.

2. Demographic Stochasticity

Changes in population density due to the births and the deaths necessarily imply the role of chance, as birth and death, but also the sex of offspring and other parameters are random processes. It follows that there is a positive probability that all of the individuals in a popu-

Box 2

Centaurea Corymbosa from the Massif de La Clape: A Species on the Brink of Extinction

Centaurea corymbosa is a cliff-dwelling plant species, endemic to the Massif de la Clape in Southern France. It apparently cannot stand competition since it is only observed on cliffs or on rocks with very few other plant species; it cannot be found either in the pinewoods or vineyards downhill, or in the garrigue on the plateau. Although the massif is about 50 km, and that cliffs suitable for the species seem to occur all over it, only six natural populations are known, within a 3 km area, and they are separated by distances from 0.3 to 2.3 km. *C. corymbosa* is an outcrossing monocarpic perennial. Seeds germinate mostly during the autumn following summer dispersal (almost no seed bank); plants then remain at the vegetative rosette stage up to at least the fourth year following germination. Flowering takes place from April to August and plants are pollinated by small hymenoptera and diptera. About 500 individuals flower every year (from 5 to 250 per population), corresponding to about 6000 to 10,000 rosettes for the whole species (including seedlings).

Using both genetical and ecological approaches, French and Spanish researchers (Colas *et al.*, 1997) carried out collaborative work to understand the factors limiting the colonization ability of *Centaurea corymbosa*.

Although the populations are at most separated by 2.3 km, population genetic structure suggests that gene flow among populations is highly restricted, as shown by the high level of differentiation: it is much larger than usually observed for other species with similar biology.

An ecological study has showed that seed dispersal distances are very short: a few tens of centimeters from the mother plant. Contrary to other species of the same family (Asteraceae), long-distance seed dispersal has never been observed, even when the wind reached 100 km/h. At maturation, seeds fall near the mother plant and the wind just pushes them on rocks before landing into a rock or a cleft. Pollen dispersal also seems to limit seed production since the fertilization rate of plants isolated by more than 4 m is about half that of plants with a close flowering neighbor. Possible colonizers of a new cliff, a few hundreds meters from the source populations, would thus have little chance to set seeds.

Successful experimental introductions on new cliffs nearby natural populations confirmed that they are suitable and were unoccuppied because of the lack of colonization ability of *C. corymbosa*.

One might wonder why dispersal ability has not evolved in this species. It appears that landscape structure selects against larger dispersal and colonizing ability. Hence, if by chance some seeds are dispersed a long way to a suitable site, germinate, and grow into adult plants, they are likely to die without producing any offspring if they do not flower simultaneously.

The species has thus become trapped in its own deleterious system. One could argue that the species looks perfectly well adapted to the rocky habitat of the Massif de La Clape. However, it is clear that its survival entirely depends on that of the six extant populations that are very small. Moreover, there is some demographic evidence that some populations might go extinct in the near future because of demographic stochasticity, Allee effect, or inbreeding depression. To prevent this species from becoming extinct, more introductions into new sites will have to be performed.

lation die during the same time interval. This probability depends mainly on the population density—the ones with lowest density being the most vulnerable—and on the mean and variance of the survival rate.

3. Environmental Stochasticity

In natural populations, the probability of having offspring or of dying are influenced by the environment, and these environmental effects have some degree of correlation among individuals. If the correlation is low, a failure, say, of one individual in producing offspring, may be canceled out, at the level of the population, by a large number of offspring produced by another individual. If, however, the correlation is high, the stochasticity will reduce the variance of the demographic parameters among individuals and therefore increase the risk of extinction. In the extreme, during a total catastrophe the correlation is complete, and all of the individuals die at the same time. Thus, high population densities are by no means a guarantee of population survival.

More formally, environmental stochasticity is evident in the temporal fluctuations of the growth rates of the populations. The increase in the variance of population growth increases the risk of extinction, in particular when mean growth rate is low, population density is low, or the carrying capacity of the environment is low.

4. Genetic Factors

Two widely discussed genetic factors associated with extinction, both of which are important in small populations, are inbreeding depression and the loss of evolutionary adaptability to new environments.

If a population goes through the bottleneck of a low population size, individuals are constrained to mate with close relatives. This is often associated with reduced viability or fertility, a phenomenon called inbreeding depression. The mechanisms leading to inbreeding depression are unknown for the majority of the species. In *Drosophila* about half of the effect is due to the increased number of homozygous loci carrying recessive detrimental alleles (up to 5000), the other half is due to the accumulation of slightly deleterious and recessive mutations. Inbreeding depression has been established in many domestic species and zoo populations, but unfortunately, information is rare for natural populations.

5. The Critical Role of Population Size

The central message that emerges from most empirical and theoretical studies on extinction is that extinction risk increases, and thus that the life expectancy of the population decreases, as population size decreases (Fig. 4). Demographic stochasticity greatly increases the extinction risk if population size decreases below about 50 individuals; inbreeding depression is likely to become significant only in circumstances when a previously abundant population goes though a population bottleneck.

B. Metapopulations: Sets of Several Local Populations

Considering a metapopulation introduces an additional dimension to the risk of extinction, due to the possibility of dispersal from one local population to another.

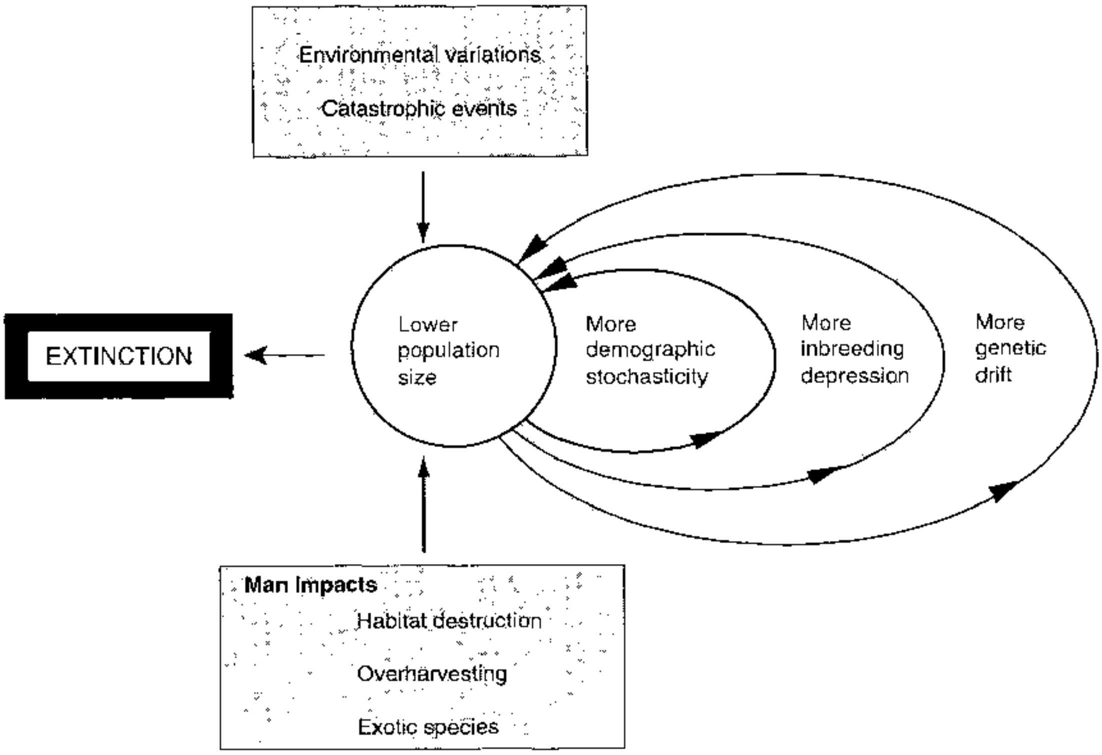


FIGURE 4 Extinction vortices: the smaller a population becomes, the more vulnerable it is to demographic, genetic, and environmental factors that tend to reduce population size even more and drive the population to extinction. Modified from Primack (1992).

Seeds, spores, or animals may disperse passively or actively from their natal population over the landscape. If they arrive at a favorable site that is not already occupied by members of their species, they may be able to found a new local population. In the long run, populations of a species may thus be established and disappear repeatedly at the local scale, while at the level of the landscape, one observes the dynamic equilibrium of a species moving among local habitats.

The persistence of a metapopulation depends on processes that affect rates of extinction of local populations, but also on processes that affect the establishment of new local populations. Two conditions must be fulfilled to get a new population established: favorable sites must exist, and a sufficient number of individuals must migrate from existing sites to the unoccupied habitats. That makes immigration and dispersal processes important mechanisms for the survival of metapopulations.

The survival of a metapopulation is threatened by two types of stochasticity, which are analogous to the demographic and environmental stochasticities involved in the dynamics of local populations: (a) the stochasticity of local colonization and extinction rates and (b) the regional stochasticity. The first involves the stochastic processes discussed earlier; the second, analogous to environmental stochasticity in isolated populations, applies if stochasticity has effects that are correlated among several local populations within the metapopulation. Since a significant source of environmental stochasticity is related to weather conditions, which are typically strongly correlated over a large spatial scale, one may generally expect a high degree of correlated regional stochasticity in metapopulations and thus high risks of extinction.

IV. CAUSES OF THE LOSS OF SPECIES AND GENES

We have argued as if there were no change in the environment experienced by the populations, but there is, of course, hardly any doubt that the current increase in the rates of extinction is primarily caused by drastic changes in the environment. This affects all of the key parameters associated with the risk of extinction: population sizes, the carrying capacity of the environment, the mean and variance of growth rates, the genetic structure of a population, the number and area of suitable habitats, and the number of local habitats forming a metapopulation. With this perspective, we will now consider three types of environmental changes that may initiate the extinction process: the destruction of natural habitats, overexploitation by humans, and changes concerning other species.

A. Destruction and Changes of Natural Habitats

All animal and plant populations are adapted to the local conditions of their environment; their persistence depends on the maintenance of these conditions. Large climatic changes of the past, the periods of glaciation, for example, led to local extinctions and shift of species' geographic ranges. These kinds of changes, however, were very slow, being spread out over periods of millions of years, so that most species were able to gradually adapt or alter their geographic range according to the novel conditions. Humankind has introduced a novel dimension to the changes of natural habitats; human activities are more similar to volcanic eruptions than to gradual climatic change and thus could amplify its effects. At the time of the last glaciation, for example, the extinctions of the large mammals during the Pleistocene are partly manmade.

The main effects of humans on the environment are the following:

1. Destruction of habitats (clearing forests, drying lakes and humid areas, etc.)
2. Degradation of natural habitats (pollution, dams, etc.)
3. Fragmentation of habitats

1. Destruction of Habitats

The most severe threat imposed on biological diversity is the loss of suitable habitat—and reducing habitat loss is the most important way of preserving the environment. The loss of habitat has been the main source of extinction, at least in the recent extinctions of vertebrates (Table IV). In many areas of the world, particularly on the islands and everywhere where the human density is high, a large portion of the natural habitats has already been destroyed (Box 3). Thus, in tropical countries, the destruction of the habitat ranges from 29% (Zambia) to 89% (Gambia) in Africa and from 41% (Malaysia) to 97% (Hong Kong) in Asia, and in most countries the number is above 70%.

Currently, the global rate of deforestation is between 1 and 2% per year, depending on the region. The main reason for such a rapid destruction of humid tropical forests is agriculture and other sources of human pressure, though the exploitation of forests for the trade of wood is also a significant cause.

TABLE IV

Factors Responsible for Some Extinctions and Threatened Extinctions[a]

	Percentage due to each cause[b]					
Group	Habitat loss	Over-exploitation	Species introductions	Predators	Other	Unknown
Extinctions						
Mammals	19	23	20	1	1	36
Birds	20	11	22	0	2	37
Reptiles	5	32	42	0	0	21
Fishes	35	4	30	0	4	48
Threatened extinctions						
Mammals	68	54	6	8	12	—
Birds	58	30	28	1	1	—
Reptiles	53	63	17	3	6	—
Amphibians	77	29	14	—	3	—
Fishes	78	12	28	—	2	—

[a] From Reid and Miller (1989).

[b] The values represent the percentage of species that may be influenced by more than one factor: thus some rows may exceed 100%.

2. Fragmentation of Habitats

The fragmentation of habitats is the process by which a large area consisting of a given type of environment is reduced in size and divided into two or more fragments. These fragments of the original environment are often isolated from one another by a highly modified and degraded landscape. For example, the large stands of vegetation that used to cover vast areas of the world have been more and more divided into separated fragments. Europe as well as densely populated areas of Asia have experienced this process for centuries.

Such isolated fragments differ from the original environment in two critical ways:

1. The ratio of edge to area is increased.
2. The center of each fragment is nearer the edge.

Fragmentation of the habitat threatens the persistence of the species associated with it in various ways.

First, fragmentation may limit the capability of a species to migrate to, and thus to colonize, other habitats, because the landscape between the fragments creates an effective barrier to dispersal. In an undisturbed habitat, individuals can disperse over the whole landscape, which can lead to a stable metapopulation despite local extinctions (discussed earlier). If, however, the habitat is fragmented, the potentials for dispersal and for colonization are often reduced. Thus, many birds or mammals, and many insects, are reluctant to leave the cover of their forest habitat and to cross small open spaces, as, if they do, they may become the victim of predators prowling forest edges. Cultivated fields, measuring only 100 m in width, can constitute an insurmountable barrier to numerous invertebrate species. If the dispersal ability of, say, some mammals is reduced, plants might also suffer, as they may rely on these species for the dispersal of their seeds. Therefore, isolated fragments of habitat may develop that, though in principle suitable for colonization, cannot be reached by dispersing species; they will therefore remain uninhabited.

As some species disappear from a fragmented landscape as a result of natural processes in a metapopulation, while others do not succeed in recolonizing the habitat, the number of species will decline over time. It must be noted here that, worldwide, most natural parks and reserves are too small and too isolated to harbor a large number of species; almost half of the protected sites have an area of less than 100 km^2, and 98% have an area less than 10,000 km^2.

By splitting a large and contiguous population into subpopulations restrained to limited areas, the fragmentation of the environment can precipitate the decline of species and their extinction. As we have seen, small populations are indeed more vulnerable to inbreeding depression, to genetic drift, and to several other problems. While a large contiguous area may support a single large population, it may well be that none of the fragments can harbor a population large enough to ensure its long-time persistence.

Box 3

Patterns of Deforestation

Wildlife habitat loss is one of the foremost threats to biodiversity and the destruction of tropical rain forests has come to be synonymous with the loss of biodiversity. In fact, tropical moist forests occupy 7% of the earth's land surface but are estimated to contain over 50% of its species (Myers, 1986).

The original extent of the tropical rain forests has been estimated at 16 million km^2, based on current patterns of rainfall and temperature. From a combination of ground surveys and remote sensing data from satellites it is possible to assess the present area of tropical forests as well as the percent deforested per year: it ranges from 0.5% per year (Brazil) up to 3% (Ivory Coast), with an avearge around 2% year.

However, estimates of deforestation rates may vary tremendously, depending on the definitions and assumptions used in quantifying both the forest's original extent and the amount of forest that has been converted to another land use.

On a global scale, the primary cause of rain forest destruction is small-scale cultivation of crops by farmers (45,000 km^2 per year). Another 45,000 km^2 per year is destroyed through commercial logging. A further 25,000 km^2 is degraded for fuelwood production, mostly to supply local villagers with wood for cooking fires. The remaining 20,000 km^2 per year is cleared for animal breeding and pasture for cattle (Table V). The relative importance of these activities varies by geographical region, with logging being a more significant activity in tropical Asia, cattle ranching being more prominent in tropical America, and farming and fuelwood gathering more important in tropical Africa.

The national figures obscure the probable impact of humans in the forest, because no allowance is made for the effects of forest fragmentation. This phenomenon has not yet been assessed at the global level.

Second, the increased proportion of habitat edges in fragmented environments brings with it changes in microclimate—light, temperature, humidity, wind—but also increased risk of fire and higher vulnerability to novel species of predators and competitors (by, for example, invasion of weedy species from disturbed environments).

B. Persecution and Exploitation of Populations

The persecution and exploitation of natural populations generally threatens a certain number of species with extinction, in particular the large-bodied ones. These, and in particular the vertebrates, are also those that attract most attention from the general public, from protection agencies, and from the scientific community. This is reflected in the considerable amount of scientific literature specializing on a rational and controlled exploitation of natural populations, be it by hunting or fishing.

One generally acknowledges that large animals require larger reserves than small ones, due to their low population densities. However, though it is generally true that large animals have small population densities, this may be due in many cases to the past persecution or exploitation of these species. It must be emphasized that large-bodied species may survive in disturbed habitats, if they are not persecuted. Thus, in Finland, large birds and mammals, including predatory species, have become more abundant recently in a range of manmade landscapes, most likely because they are now protected.

C. Changes in the Biotic Environment

A common cause of extinction is the interaction with exotic species that had either been introduced by humans or arrived naturally (Box 4). The introduction of exotic species threatens in particular small isolated islands with specialized endemic species. Large islands, however, are not protected from this effect: after being introduced to Australia, the red fox became the major cause of extinction of several small marsupials.

The isolation of insular habitats favours the evolution of endemic species, but it also makes these particularly vulnerable to invading exotic species. Only a limited number of species can reach islands; plants, birds, and invertebrates are the most common colonizers. On the other hand, insular communities generally have few or no predators and browsers (due to the difficulties of colonizing the island or of establishing a population on an insufficient area), and indeed, the species representing the highest trophic levels (e.g., the carnivorous mammals) may be missing completely. Since many endemic species on islands have thus evolved in the absence of the selective pressures of predators and mam-

Box 4

Examples of Extinction Cascades due to Introduced Species

What happened to Guam's avifauna is a good example of the destructive effects of introduced species. Guam is an island belonging to the Marianes Archipelago, situated between Japan and New Guinea. Eighteen species of indigenous bird species were known to inhabit the island, as well as seven introduced ones. Then, during the past 20 years, populations declined spectacularly: today seven species are considered to be extinct and four others have become so rare that their survival is endangered. Nevertheless, on other islands of the same archipelago, no comparable decline has been observed. Strikingly, the 10 forest species were all affected and in a similar manner: the birds disappeared first from the southerly forests during the 1960s, and then their decline spread progressively toward the north of the island. This decline coincided exactly with the introduction and subsequent spatial expansion of a tree snake, *Boiga irregularis*. Since this snake is absent on neighboring islands, there is good evidence that it was the cause of the observed declines and extinctions. Because of its arboreal and nocturnal habits, this snake is a voracious predator of perching birds and birds sitting on their nests, as well as of eggs and nestlings. Moreover, because it predates on small mammals and lizards as well (the latter particularly abundant), it can attain high densities even while exterminating its most vulnerable prey species.

On Santa Catalina, an island close to California, 48 native plant species have been eliminated mainly as a consequence of overgrazing by goats and other introducted mammals.

At Madagascar, where the ichthyofauna is highly endemic, with 14 out of 25 genera unknown elsewhere, a recent inventory of freshwater environments was not able to retrace more than 5. Introduced fishes dominate all aquatic environments. The combination of habitat degradation and the introduction of exotic fishes seem to lead the original ichthyofauna toward extinction.

The same has been found in continental aquatic environments, which for aquatic species are a kind of island surrounded by inhospitable terrestrial space. Originally, Lake Victoria had more than 350 endemic fish species. Today, many are rare or have gone extinct after the introduction of the Nile perch, *Lates nilotica*, in 1960. In fact, the phenomenon is more complicated than that: in 1978, the perch did not yet represent more than 2% of the annual catch; in 1986 it represented 80%. Other factors than predation have played a role (algal blooms creating anaerobic conditions as a consequence of pollution by fertilizers and other pollutants).

malian browsers, they have not evolved (or have lost) any means of defense: birds have lost the ability to fly and lay their nests on open ground; plants do not produce any chemical substances and do not have any protective tissues (spines) that could deter browsers.

Therefore, endemic species that can spread in the absence of these selective pressures may go extinct rapidly when the pressures appear: animals introduced into the islands eliminate them by predation or overgrazing. The introduced plants, however, equipped with a protected or toxic foliage, have previously evolved to withstand browsing, so that they outcompete the endemic plants, enhancing the selective pressure by the introduced browsers and thus accelerating the extinction of the endemic species.

Moreover, insular species usually do not have an immune defense against foreign diseases; once introduced with the invading animals, these can spread epidemically and can devastate the native populations. It is thought, for example, that the almost complete destruction of the Hawaiian avifauna—since 1850, 85% of the endemic species have gone extinct or have been reduced to very small populations—is due to the spread of variola and of bird malaria, introduced together with the mosquito *Culex quinquefasciatus* in 1826.

Manmade extinctions in insular communities are much more frequent than generally recognized. Thus, the rate of extinction on the order of 1% given earlier is a gross underestimate of the true value. Current estimates are that 25% of endemic bird species have disappeared from islands as a result of human activities. There are, for example, only three species of winged rails today, while thousands have existed during the past 2000 years. We are rightly worried by the current wave of mass extinctions threatening the tropical forests, but should also recognize that the oceanic islands have already experienced such mass extinctions.

D. Ecological Reflection on the Primary Causes of the Current Crisis of Extinction

All organisms modify their environment, and humans are no exception. Thus, in a strategic document, jointly published by the World Resources Institute (WRI), the World Conservation Union (IUCN), and the United Nations Environmental Programme (UNEP), six fundamental causes for the impoverishment of biodiversity are recognized:

1. The rapid and unsustainable growth of human populations and the consumption of natural resources
2. The continued reduction of the range of products in agriculture, forestry, and fishing
3. The economic and political systems that do not take into account the environment and its resources
4. The inequalities in the possession, the management, and the sharing of the advantages related to the use and the conservation of biological resources
5. The legislative and institutional systems that favor unsustainable exploitation of resources
6. The lack of knowledge and its applications.

In fact, diversity decline results from specific choices of path during the course of human development (Swanson, 1995). It is clear that, with population growth, demographic changes, and the associated technological development, humankind is using an increasing share of the planetary resources. Indeed, it consumes, diverts, or holds approximately 39% of plant productivity, the fundamental source of energy for most living systems (Vitousek *et al.*, 1986). The rate of conversion from natural environments into agricultural systems, high in the developing countries (Table V), remains a major threat for biodiversity. This is alarming not only for the future of many animal and plant species, but also for our own.

TABLE V

Recent Rates of Conversion of Natural Habitat to Specialized Agriculture (the 10-year rate, up to 1987)[a]

Conversions to cropland (%)		Conversions to pastureland (%)	
Paraguay	71.2	Ecuador	61.5
Niger	32.0	Costa Rica	34.1
Mongolia	31.9	Thailand	32.1
Brazil	22.7	Philippines	26.2
Ivory Coast	22.4	Paraguay	26.0
Uganda	21.4	Viet Nam	14.0
Thailand	17.1	Nicaragua	11.8

[a] Source: World Resources Institute and International Institute for Environment and Development.

In addition to the dangerous reduction of the renewable resources from the environment, population growth leads to increases in pollution (including the gases responsible for the greenhouse effect), which threaten the balance of ecosystems and of our planet's climate. Novel models of development are necessary so that population growth does not exceed the carrying capacity of the planet.

The five other primary causes blame either our economic behaviors or the legal uses of the human societies. For thousands of years, the human world formed a mosaic of relatively autonomous areas. The state of knowledge, the strategies of subsistence, and the social structures evolved more or less independently in each area. What the populations required of the environment seldom exceeded the capacity of nature.

With the globalization of the economy, starting at the end of past century, the uniformity and the interdependence have increased. In agriculture, for example, a small number of cultivated plants have come to dominate the global economy (Box 5). In fact, the human societies today mainly depend on four species of plants (maize, wheat, potato, and rice) for their basic needs. Furthermore, these species are used in the form of a decreasing number of high-output varieties. Thus, on Sri Lanka, the number of 2000 varieties of rice used in 1959 has decreased to 5 main varieties today; in India 10 varieties out of originally 30,000 represent 75% of production; about 62% of the varieties in Bangladesh and 74% in Indonesia are estimated to be derived from a single maternal plant (Box 5).

The process of "conversion," which consists in investing capital in a type of resource that is economically advantageous, is the principal economical force threatening biodiversity: "Development and conversion go hand in hand, and conversion is the process by which habitat and its resident species are lost" (Swanson, 1995).

The reduction of the number of cultivated species is accompanied by the disappearance of nitrogen-fixing bacteria, mycorrhiza, predators, pollinators, and many other species that have coevolved during centuries with the traditional systems of agriculture. The use of ma-

Box 5

Modern Agriculture: A Biodiversity-Eroding Factor

It is rather surprising to note that human civilization uses a severely reduced number of species for agriculture and livestock: a couple of hundred plant species and a couple of tens of animal species. Most of the agricultural production is represented by at most 20-odd of these.

Almost all of present-day agronomical research, whose aim is to increase the yield of agriculture, focuses on the improvement of cultivars of the 20 main plant species that provide about 80% of the annual world harvest. Four of those species—wheat, maize, rice, and potato—already account for almost half of it.

Furthermore, the Green Revolution has led to the adoption of modern varieties and at the same time a considerable genetic erosion. Thus, in Indonesia, 1500 local varieties of rice have disappeared during the past 15 years. In Kenya, of 13 wild coffee varieties, only two are not endangered and two have already disappeared. In 1991, the genetic homogeneity of orange trees in Brazil has favored the worst epidemic of root cancers seen in the country. In Asia, following the Green Revolution, a considerable proportion of rice varieties originate from a single mother-plant (62% in Bangladesh, 74% in Indonesia, 75% at Sri Lanka): similar epidemics can emerge at any moment.

Such an economic concentration on such a low number of species is even more paradoxical when it is taken into account that over 3000 plant species are edible! Worse even, during the past couple of centuries, species known and cultivated by indigenous populations have seen their use reduced, a consequence of cultural homogenization due to colonization.

One of the most well-known cases is that of the amaranth (three species of the genus *Amaranthus*), of which the Aztecs consumed the seeds and whose culture was forbidden by the Spanish when they conquered Mexico. The same is true for a plant of the high Andean plateaux, the passerage (*Lepidium meyenii*), whose roots resemble those of the black radish and are rich in saccharose and starch. These plants, which used to be grown by the Incas of Peru and Bolivia, are on their way to extinction as they no longer cover more than about 10 ha.

nure, pesticides, and high-output varieties accelerated this erosion.

Similarly, with regard to marine ecosystems, the vast world markets favored the development of "blind" fishing with, for example, gigantic drifting nets that trap not only enormous quantities of target species, but also a considerable number of other species of fish, mammals, and birds.

In addition, several reasons led to an underestimate of the use of environmental resources. First, many resources are directly consumed without appearing on the markets. Second, the benefits of biodiversity are largely public goods. For example, while the protection of wetlands may benefit a population, the benefit is so diffuse that traditional economic theory cannot incorporate mechanisms that would preserve them. Correctly evaluated, the natural systems and their biodiversity are major economic assets. But as these systems are often underestimated, economic theory considers the conservation of the biodiversity an expenditure rather than an investment.

In response to this challenge, a new discipline is being developed that integrates economics, ecology, and public policy: ecological economics. As Thimothy Swanson emphasized, "diversity decline is a specific form of institutional failure: the failure to create institutions that internalise the values of biodiversity within the decision-making of states and individuals making conversion decisions."

The reduction in the number of the species and the destruction of the habitats are the standard in many countries where a minority of the population has or controls most of the territory. The rapid profits obtained from excessive logging or fishing go to a minority, while the local population, which depends on the sustainable production of the resources, pays the price. In fact, the ownership is often more likely to be granted to those that cut down and colonize forests and other areas covered with natural vegetation than to the inhabitants of the forests living on the sustainable harvest of the natural products. Any uncertainty on the ownership dissuades good management practices and encourages overexploitation.

A second problem stems from the concentration of the control of the resources and to the responsibilities for decisions about environmental policy in the hands of townsmen. However, in many societies, it is the women that manage the environment and understand the value of the biodiversity for agriculture and health.

A third problem is that of international trade, the debt and the policies of technology transfer that support the inequalities and often reinforce the imbalances observed within the nations. To preserve the biodiversity, the industrialized countries must reverse this flow. If the developing countries continue to be excluded from the markets, to be deprived of accesses to technologies, and to be blocked by their debts, they will have neither the means nor the incentives necessary to preserve their biological resources.

See Also the Following Articles

EXTINCTION, CAUSES OF • EXTINCTION, RATES OF • HUMAN IMPACT ON BIODIVERSITY, OVERVIEW • MASS EXTINCTIONS, NOTABLE EXAMPLES OF • METAPOPULATIONS

Bibliography

Barbault, R. (1994). *Des Baleines, des Bactéries et des Hommes.* Odile Jacob, Paris..

Barbault, R., and Sastraprajda, S. (1995). Generation, maintenance and loss of Biodiversity. In *Global Biodiversity Assessment* (V. Heywood and I. Baste, Eds), pp. 193–274. Cambridge University Press, Cambridge.

Blondel, J. (1995). *Biogéographie. Approche écologique et évolutive.* Masson, Paris.

Colas, B., Olivieri, I., and Riba, M. (1997). *Centaurea corymbosa: a cliff-dwelling species tottering on the brinks of extinction.* Proceedings of the National Academy of Sciences, USA, 94, 3471–3476.

Dobson, A. P. (1995). *Conservation and Biodiversity.* Scientific American, New York.

Laurence, W. F., and Bierregaard, R. O., Jr. (1997). *Tropical Forest Remnants.* Chicago University Press, Chicago.

Lawton, J. H., and May, R. M. (Eds.) (1995). *Extinction Rates.* Oxford University Press, Oxford.

Meffe, G. K., and Carroll, R. C. (1997). *Principles of conservation biology*, 2nd ed. Sinauer, Sunderlands.

Primack, R. B. (1993). *Essentials of Conservation Biology.* Sinauer, Sunderlands.

Raup, D. M. (1991). *Extinction: Bad Genes or Bad Luck?* W. W. Norton and Company, New York.

Reid, W. V., and K. R. Miller (1989). *Keeping Options Alive: The Scientific Basis for Conservation Biodiversity.* World Resources Institute, Washington, D.C.

Sepkoski, J. J., Jr. (1992). Phylogenetic and ecologic patterns in the Phanerozoic history of marine biodiversity. In *Systematics, Ecology and the Biodiversity Crises* (N. Eldredge, Ed.). Columbia University Press, New York.

Smith, F. D. M., May, R. M., Pellew, R., Jonshon, T. H., and Walter, K. S. (1993). Estimating extinction rate? *Nature* 364, 494–496.

Swanson, T. (1997). *Global Action for Biodiversity.* Earthscan Publications, London.

Vitousek, P. M., Ehrlich, A. H., and Matson, P. A. (1986). Human appropriation of the products of photosynthesis. *Bioscience* 36, 368–373.

Wilson, E. O. (1992). *The Diversity of Life.* Allen Lane, The Penguin Press, London.

World Resources Institute (1992). *Global Biodiversity Strategy.* World Resources Institute (WRI), the World Conservation Union (IUCN), and the United Nations Environmental Programme (UNEP).

MAMMALS, BIODIVERSITY OF

Joshua R. Ginsberg
Wildlife Conservation Society

- I. Introduction to Mammals
- II. Mammalian Phylogeny
- III. Phylogeny and Biodiversity: Patterns of Ordinal Diversification
- IV. Evolutionary Trends
- V. Geography, Biogeography, and Biodiversity
- VI. New Discoveries
- VII. Current Extinction Crisis

GLOSSARY

brachyodont Low-crowned cheek tooth commonly found in omnivorous or burrowing animals.

cloaca The chamber into which the gut and urinary and reproductive tubes empty before exiting the body.

craniomandibular Where the head, or cranium, and the jaw, or mandible, meet.

cryptic species Species that are virtually indistinguishable by normal morphological analysis and are, instead, defined by a combination of genetic, behavioral, and other characters.

cursorial Adapted for running.

diastema A large gap between incisors and premolars that is thought to permit an animal space to manipulate food with the tongue.

digitigrade Literally, walking on the digits; a posture where the majority of the weight is borne by the metacarpal and metatarsal bones and the heel/palm is raised off the ground.

endothermy The metabolic condition of being hot-blooded and internally regulating body temperature.

eutherian A placental mammal of the subclass Eutheria.

fossorial A burrowing animal.

hypsodont High-crowned cheek tooth commonly found in grazing mammals.

keystone species A species that exerts an effect on the structure and function of a community out of proportion to its numerical abundance.

macropod A marsupial in the family Macropodidae, the kangaroos and wallabies.

marsupium Folds of skin that envelope the mammary glands and provide protection to infants; found in metatherian and protherian mammals.

metatherian A member of the infraclass of mammals that includes the marsupials.

monotreme A mammal in the family Monotremata, which includes spiny anteaters and the platypus.

monophyletic Descriptive of a clade, or branch, of an evolutionary tree that has a single root, indicating that all members of that branch are descended from a common ancestor.

patagium The skin that stretches across the arms and legs of flying and gliding mammals.

phylogeography The study of the relationship between and among genetics, morphology, paleontology, and ecology to better understand the spatial aspects of evolution.

placental mammal See *eutherian*.

polyphyletic Descriptive of a group of animals that, although grouped together in a phylogeny or evolutionary tree, do not share a common ancestor.

ricochetal Bouncing movement that involves the release of stored energy in stretched tendons, movement that involves quick changes of direction, as in *to ricochet*.

rumen The first, and usually largest, chamber of the four-chambered stomach found in ruminant ungulates; the place where cellulose digestion occurs.

semelparous Species that breed once and die.

sympatric Populations or species that co-occur in a single location.

RANGING IN SIZE FROM A SHREW, which may weigh only a few grams, to the blue whale, which weighs over 150 metric tons, mammals are found in the air, on the ground, under the earth, and in the oceans, rivers, and lakes of the world. Modern mammals are represented by 135 families divided into 1135 genera, not an impressively diverse group of animals. Throughout geological history, there have been just over 400 families of mammals and perhaps as few as 5000 genera. Paleontologically, the structure of the mammalian middle ear is unique, with three small bones that were originally found in the jaw of more primitive vertebrates now forming the structure of the middle ear. Mammals are endothermic; i.e., they regulate their internal temperature. To support endothermy, mammals have greater cellular production of energy, lungs with a large capacity, and numerous red blood cells to transport oxygen; an efficient four-chambered heart; and increased digestive efficiency. Soft-tissue characters that are diagnostic of mammals include sweat glands, hair, a folded cerebellum, an epiglottis, and a complex lung structure. Of course, mammary glands are the single most common diagnostic feature for mammals: the production of milk with which to feed and nurture young is a shared feature of all mammals.

I. INTRODUCTION TO MAMMALS

Modern mammals are not an impressively diverse group of animals, at least not numerically. Mammals are an extremely well described class of animals, with, perhaps, only Aves being better examined. Well described and relatively well studied, modern mammalian diversity is limited to 26 orders, divided into 135 families with 1135 genera and classified, at present, into approximately 4650 species (Fig. 1). Mammals have about half

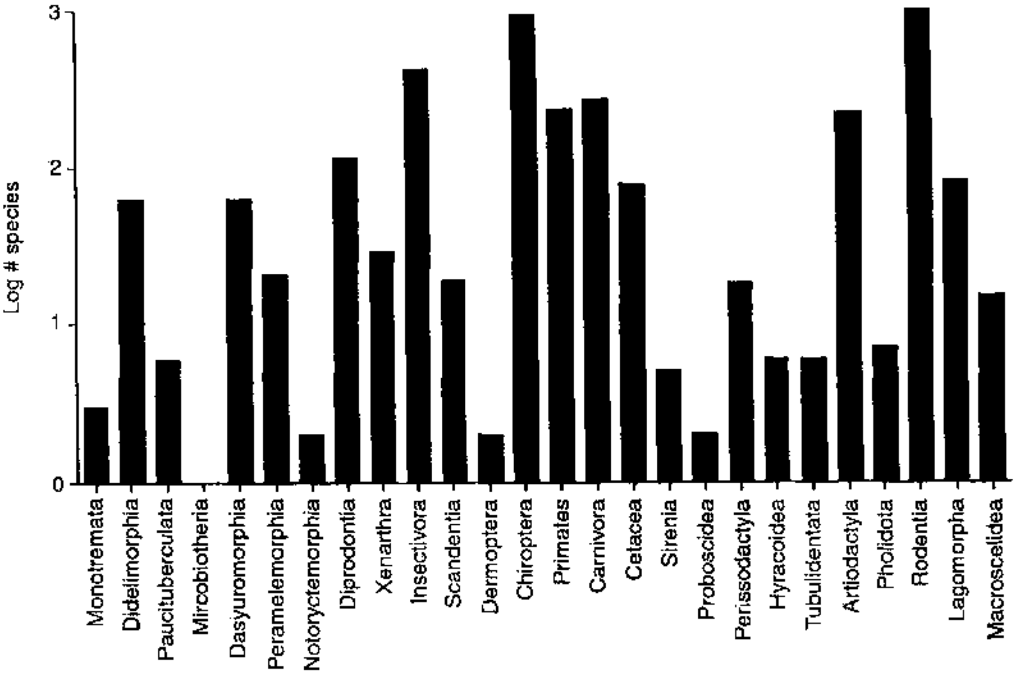


FIGURE 1 Species diversity of recent mammalian orders. Data derived from Vaughan *et al.* (2000). All taxonomy is based on Wilson and Reeder (1999) as modified by Vaughan *et al.* (2000).

the diversity of modern birds and pale in significance when compared to the nearly million described species of insects and the perhaps 10 million species that remain undescribed. Historical consideration does little to mediate these discrepancies in diversity: throughout all geological history, there have been just over 400 families of mammals and perhaps 5000 genera.

What mammals lack in numeric diversity is made up in their physical and ecological variation. Ranging in size from just a few grams (e.g., shrews in the family Soricidae) to over 150 metric tons (the blue whale, *Balaenoptera musculus*), mammals are found in the air, on the ground, under the earth, and in the oceans, rivers, and lakes of the world. Their diversity of body types, diets, and habits has captivated both the general public and scientists, taking an entirely disproportionate share of scientific study and popular examination.

A. What Is a Mammal?

There are literally dozens of diagnostic characteristics of mammals, most of which are related to soft-tissue structure and metabolic function. For paleontological purposes, soft-tissue differentiation of mammals is of little or no value. Soft tissue is rarely preserved and its structure can only occasionally be deduced from the structure of the bones that it covers and from which it is suspended. Paleontologists, therefore, look for skeletal and cranial features that distinguish mammals from other classes. One defining character is found in the structure of the mammalian middle ear. The middle ear contains three small bones, the stapes, incus, and malleus, which were originally found in the jaw in the craniomandibular joint of more primitive vertebrates (the joint where the jaw meets the cranium). The simplification of the craniomandibular is also reflected in further simplification and solidification of cranial bones overall. This simplification and increased ossification are also reflected in the postcranial skeletal structure of mammals. Skeletal growth occurs early in the life history of mammals, with further growth of bones limited to a cartilaginous area at the end of the bone, under the articular surface, called the epiphysis. Eventual closure and ossification of the epiphyses occur after growth is complete.

Physiologically, a critical difference between mammals and reptiles is the development of endothermy, or internal regulation of temperature. While it is unclear whether the mammal-like Therapsids were endothermic, all modern mammals exhibit endothermy, while extant reptiles are ectothermic, using behavioral adaptations to regulate temperature in harsh environments or allowing their temperature to fluctuate widely. The cellular and structural differences that endothermy require are all hallmarks of mammalian life: greater cellular production of energy, lungs with a large capacity and numerous red blood cells to transport oxygen, efficient circulatory systems, and increased digestive efficiency. Soft-tissue characters that are diagnostic of mammals (many of which are related to endothermy) include sweat glands, hair, a four-chambered heart, a folded cerebellum, an epiglottis, and a complex lung structure. Of course, mammary glands are the single most common diagnostic feature for mammals: the production of milk with which to feed and nurture young is a shared feature of all mammals.

B. Overview of the Class Mammalia

Mammals are usually divided into three subclasses: the Allotheria, which was represented by the extinct multituberculates; the Prototheria, which reached the peak of its diversity in the Mesozoic and which today is represented by a single order, Monotremata, containing the two species of echidna (Tachyglossidae) and the duck-billed platypus (*Orithorhynchus anatinus*); and the Theria, which contains two infraclasses, the Metatheria (marsupials) and the Eutheria (placental mammals).

1. Monotremata

Viviparity, or live birth of young, is so common among mammals that it is usually, wrongly, considered a defining character of the class. The three species of the order Monotremata all lay eggs. The platypus lays its eggs into a nest, similar to a bird's nest, while both species of the family Tachyglossidae, the echidna, or spiny anteaters, lay their eggs directly into a marsupial-like pouch. While this curiosity is the root of the common name for the monotremes—egg-laying mammals—the egg is actually a rather insignificant aspect of the monotreme's life history. Incubation is brief, under 2 weeks, following which monotreme development does not differ significantly from that of other mammals. Clearly mammals, they nurture their young with milk that is expressed from mammary glands that lack nipples. Like all mammals, they are endothermic, have hair, possess a single jaw bone, and have the diagnostic three-bone middle ear structure.

Divergence of the monotremes from other mammals occurred approximately 175 million years ago early in mammalian history. Fossil monotremes have only been found from Australasia, and all extant species share this distribution. Monotremes appear to be extremely primitive in their reproductive habits, with not only an egg-laying habit but also a single opening, or cloaca, into which both the excretory and reproductive tracts exit. The cloaca (or single exit) gives the order its name.

2. Theria

The therian mammals all share the derived character of live birth of their young. The subclass Theria is divided into two infraclasses: metatherians and eutherians. Metatherians are usually called marsupials because of the pouch, or "marsupium," that many members of the subclass possess. A far more diverse order than Monotremata, there are 273 described species of marsupials, which are categorized into 19 families. Metatherians and eutherian mammals probably diverged at about the same time as the monotremes, approximately 100 million years B.P. Since that time, the two major groups have had a divergent, and sometimes convergent, evolutionary history.

3. Metatheria: Marsupials

In Recent times, metatherian mammals have only been found in the Australian subcontinent, the adjacent large islands of New Guinea (the island consisting of Papua New Guinea and the Indonesian state of Iryan Jaya) and Tasmania, and the Neotropics. This distribution is more restricted than that given by fossil evidence. Arising in the Mesozoic, about 100 million years ago, marsupials are first found in North and South America. The metatherians make their first appearance in the fossil beds of the Eocene of Europe, 50 million years ago, and it is thought they arrived in Asia, and Australasia, some time thereafter. The fossil evidence for this period from these regions is poor, and some have suggested that this, combined with the present distribution of marsupials, strongly implies an Antarctic origin.

In the New World, the Recent marsupial fauna has not developed significant diversity in competition with placentals. The New World family Didelphidae is represented by 63 Recent species in 15 genera. While found across North, Central, and South America, there is relatively little variation in size, with the smallest member of the family the size of a mouse and the largest only the size of a large cat. The most common species in the Americas is the Virginia, or common, opossum (*Didelphis virginiana*), which is widespread and extremely successful in adapting to human presence. The earliest members of this genus are found in South America and it appears that they migrated to North America just over a million years ago after the establishment of the Central American land bridge. While marsupials, for the most part, did poorly when they invaded North America, *Didelphis* is a clear exception to this rule.

In South America during the late Miocene and Pliocene epochs (approximately 5–10 million years B.P.), a diverse radiation of marsupials occurred. Convergent evolution produced a suite of marsupials that resembled contemporaneous placental mammals and included forms similar to large felids (in the family Thylacoleonidae). The most diverse family of carnivorous South American marsupials was the Borhyaenidae, often called the family of dog-like marsupials because of the general resemblance of some species to canids (e.g., *Lycopsis*). Borhyaenids persisted only until the late Pliocene. Other families of South American fossil marsupials included rodent-like, mole-like, and rhino-like animals and perhaps the best known, the Thylacosmilidae, or the "false" saber-tooth tigers (*Thylacosmilus*).

In the Australian region, until Recent times, the metatherians have been free from competition with placental mammals, with the exception of bats and some muroid rodents (rats and mice). Here, the metatheria have undergone a remarkable and fascinating radiation. Many species resemble, in basic form, the eutherian moles, squirrels, and small carnivores. These species tend to be relatively undifferentiated and follow a relatively similar body form within groups. In some cases, however, marsupials in Australia have evolved into forms that little resemble in structure their ecological replacements in areas where placental mammals dominate. While most grazing placental mammals are, to a greater or lesser extent, relatively similar in their body plan, the kangaroos and wallabies (Macropodidae) represent a distinctly different solution to the design of a grazer. The origin of the large ricochetal hind feet and vestigial, but highly flexible, forearms suggests an arboreal origin for the family.

4. Eutheria: Placental Mammals

The greatest diversity of mammals is seen in the eutherians. Both through geological time and in the present, eutherians are the most diverse of the three main branches of mammalian evolution. Representing nearly 95% of the extant species of mammals, eutherians have shown a remarkable variety of forms and adaptations. Fully aquatic forms are represented by the 10 families and 78 species of cetaceans. The 17 families of Chiroptera, representing over 900 species, have a near monopoly on the nocturnal skies and are keystone pollinators in tropical forests. But by far the most diverse order is Rodentia, with 29 mammalian families (approximately 25% of the total) and over 2000 species, nearly half the known mammals alive today. A more thorough discussion of eutherian diversity is given below.

One would expect that the defining character for eutherian mammals is the well-developed placenta, from which the infraclass derives its name. The use of the term is unfortunate as marsupials possess one of four types of more primitive, but nonetheless well-developed, placenta, and placenta-like structures are seen in some fish and reptiles. In all mammals, the placenta serves to provide exchange of nutrients and waste products and acts as a respiratory organ for the developing fetus. Increased complexity of the placenta found in eutherian mammals allows for more efficient transfer and for added functions, including endocrine production, including progesterone, which helps sustain pregnancy.

C. Differentiating between Placental and Marsupial Mammals

1. Cranial and Skeletal Differences

Differentiation of marsupial and placental mammals is easiest in the examination of soft tissues; nonetheless there are a set of diagnostic characters that distinguish the metatherian and eutherian mammals. While the palates of placental mammals are smooth, those of marsupials have a series of diagnostic holes or vacuities. Marsupials also differ in their dental morphology: in contrast to the serial replacement of deciduous teeth with permanent teeth found in placentals, only a single tooth is replaced in each jaw. With the exception of the third premolar, all premolar, canine, and incisor deciduous teeth are resorbed before they erupt in the jaw. Only a single set of molars forms and molars erupt sequentially. Recent fossil evidence from the Mesozoic marsupial *Alphadon* shows that this pattern of tooth replacement and development is a derived character and suggests that the trait is ancestral to marsupials.

A further primitive character common to the Monotremata and Metatheria, but thought to be lost in the eutherian mammals, is the epipubic bones, which extend forward from the pubic bone in both sexes. Hypotheses about the function of the epipubic bones have been various. Initial speculation was that the structure provided support for the attached young found in marsupials or support for the marsupium, in which they are often contained. Others have suggested it was involved with locomotion, while more recently it has been proposed that epipubic bones provided a place for the attachment of abdominal muscles, thus allowing for greater expansion of the abdominal cavity, increasing the function of the diaphragm and improving efficiency of respiration. Recent paleontological finds show that epipubic bones were present in Cretaceous eutherian mammals. Because the loss of epipubic bones is associated with prolonged gestation, these finds have been interpreted as evidence that the complex suite of adaptations that typify placental mammals were derived later in the evolution of eutherians.

2. Reproduction

While placental mammals tend to have long gestation periods and relatively short periods of lactation, gestation in the marsupials is always relatively brief and is followed by a much longer period of lactation. Typically, neonatal marsupials are extremely small, only a few grams, and show very incomplete development. The exception to this generality is seen in the jaw and forearm structure of the neonate: the nearly embryonic young need strong arms to navigate up the mother's midline to the teat and well-developed jaws with which to form an immovable attachment to the teat once arriving. A greater proportion of development occurs outside the mother's uterus after the neonate has attached to a teat. In approximately half the marsupial species, the teats are found inside a marsupium. By the time the neonate leaves the pouch, or detaches from the teat, it weighs more or less the same as a placental mammal of an equivalent adult size.

Diagnostic differences in the structure of the reproductive tract of female marsupials and in the development of the embryo may explain many of the life history differences between eutherian and metatherian mammals. While eutherian mammals form a complex placenta that nourishes the embryo, marsupials retain an eggshell membrane with a simple yolk sac to nourish the young. This arrangement limits the nutrients that a mother can transfer to her embryos.

The size of a neonatal marsupial is also limited by the structure of the female reproductive tract. In eutherian mammals, the reproductive tract is arranged linearly (Fig. 2): the vagina leads directly into the uterus, with

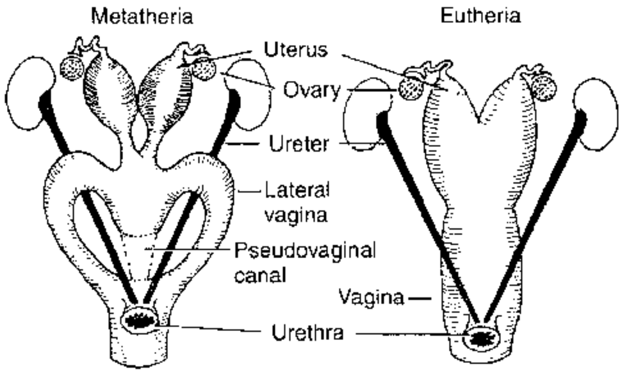


FIGURE 2 Comparison of eutharian and metatherian reproductive tracts. Reprinted with permission from Sharman, Reproductive physiology of marsupials © (1970) American Association for the Advancement of Science.

two ureters transporting urine from the bladder to the urethra. Growth of the fetus is essentially constrained only by the size and elasticity of the vaginal canal. Given that the spotted hyena gives birth through a vagina that has evolved to look like a pseudopenis, this constraint does not appear to be too great. In contrast, in marsupials, the structure of the vagina and uterus is like a double jug handle (Fig. 2). Just above the urethra, two lateral vaginas form loops on either side of a pseudovaginal canal. The two vaginal loops rejoin and feed into a bipartite uterus. The ureters pass inside these vaginal loops. Because birth takes place through the lateral vagina, placement of the ureter inside this loop may limit the size of the neonate.

II. MAMMALIAN PHYLOGENY

A. Early History

Mammals have their origins deep in geological history. Approximately 200 million years ago, in the late Triassic period, primitive cynodont mammals evolved from their mammal-like reptile progenitors, the Therapsids. The Therapsids, whose members dominated the terrestrial landscape during the Triassic, faded into insignificance by the end of the Triassic/Jurassic boundary, leaving the stage of the late Mesozoic era open to the radiation of great sauropods, the dinosaurs. For 140 million years, dinosaurs filled most terrestrial (and many aerial) ecological roles, with mammals for the most part relegated to small, terrestrial rodentlike forms.

The earliest evidence of true mammals occurs in the late Triassic, although by this time, mammals were found worldwide, suggesting a somewhat earlier divergence from the mammal-like cynodonts. Poor fossil records from many areas leave gaps, which slowly are being filled. While the Mesozoic mammals were once thought to be lacking diversity, recent finds suggest this was not the case. Fossil evidence suggests they were never abundant and rarely showed any great size—the first mammal weighing more than 1 kg does not appear in the fossil record until the early Cretaceous. Yet a variety of unusual forms evolved in the Mesozoic, including Symetrodonta, which are characterized by well-developed, triangular molars, the predatory Triconadonta, and the omnivorous and herbivorous members of the Multituberculata.

The great radiation of modern mammals began in the mid-Cenozoic era, in the late Cretaceous period, approximately 100 million years ago. More rapid evolution of larger body sized mammals did not begin until the mid-Cenozoic, beginning about 65 million years ago during the Paleocene epoch of the Tertiary. Following this, mammalian radiation accelerated.

B. Ordinal Diversification

1. Structure of the Mammalian Evolutionary Tree

The ordinal diversification of the mammals was more or less completed in the late Eocene, 50 million years ago, and has been relatively stable since that time. Of the 32 major orders that were present 50 million years ago, 27 are extant (Fig. 3). Reconstruction of evolutionary paths of ordinal diversification can be made using morphological analysis of modern or fossil taxa, molecular analysis (limited to species from which DNA can be extracted), or some combination thereof. Extant species have often lost defining skeletal and cranial characters present in earlier fossil forms but have the advantage of soft tissue being available for study. Fossils offer a long time frame, but significant gaps in the fossil record and the absence of soft tissue for comparison may be problematic. Molecular genetic studies have become increasingly sophisticated but suffer from their own inaccuracies and, of course, are useless for independently reconstructing phylogenies from taxa that have gone extinct, leaving no modern (or DNA-bearing subfossil) forms.

Morphological and molecular data agree remarkably well on the general structure of the phylogeny of the major mammalian orders, although significant questions remain about the exact relationship of several orders. While there was hope that the addition of molecular data to the discussion of mammal phylogeny would clarify some of the more confusing aspects of ordinal diversification, present molecular methods have not met that expectation.

A good example of the problem is the placement of the Hyracoidea, the hyraxes. Both protein sequence data and morphological studies suggest a close affiliation of the hyraxes with the Proboscidea (elephants) and the Sirenia (dugongs and manatees). Yet other morphological analysis suggests that hyraxes and perissodactyls (rhinos, tapirs, and equids) share a common lineage. Gene sequence data, while not without problems of its own, may offer some resolution of such problems.

2. Timing of Ordinal Diversification

While scientists agree that there was a period of rapid diversification of higher orders of mammals, the exact

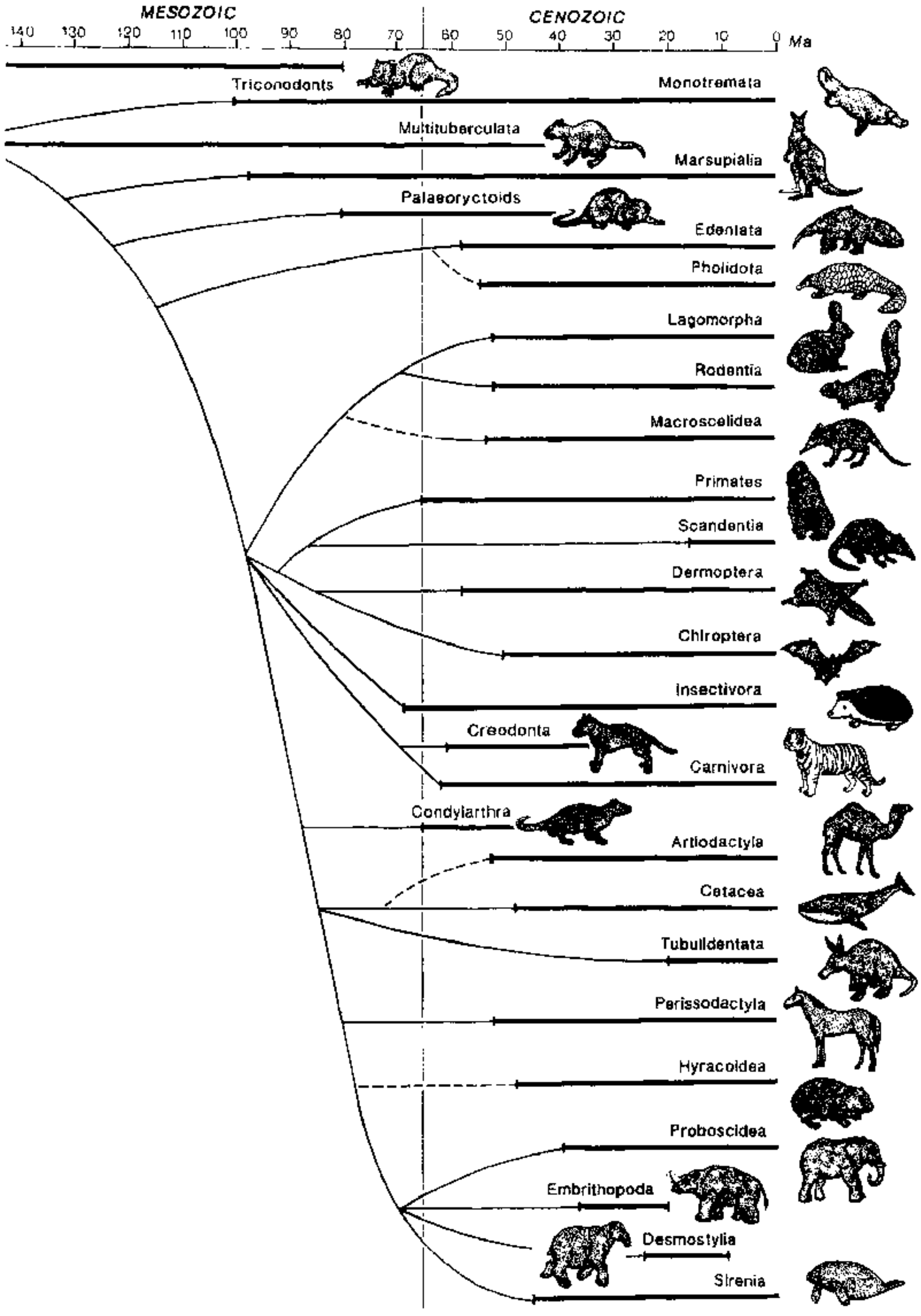


FIGURE 3 The 27 major extant orders. Reprinted by permission from *Nature*, Novacek, Mammalian phylogeny: Shaking the tree, copyright (1992) Macmillan Magazines Ltd.

timing of this event is still hotly debated. At the root of this debate are the differences in timing derived from molecular and fossil evidence. The data supporting the timing of events presented above, with the major ordinal diversification completed in the early Tertiary, are derived from the near lack of any fossil evidence for the existence of modern orders of mammals before the Cretaceous/Tertiary boundary 65 million years ago.

Advances in molecular biology have permitted more sophisticated molecular analyses of ordinal diversification of both birds (which are purported to have a similar timing of ordinal diversification) and mammals. These analyses, which employ a number of different genetic "molecular clocks," all point to an early ordinal diversification of birds and mammals some time in the middle to late Cretaceous, approximately 90–100 million years B.P. That the estimated timing of diversification is similar for birds and mammals supports this interpretation. However, of greater interest is the coincidence of the timing of ordinal divergence and the fragmentation of emergent land areas at about the same time.

By the early Cretaceous, the two main land masses formed by the breakup of Pangea, Gondwanaland, and Laurasia had further subdivided into seven continents. By the late Cretaceous, 70 million years ago, 11 continents had appeared, the peak of land mass isolation and diversity. Earlier explanations had suggested that the rapid ordinal diversity resulted as mammals (and birds) radiated into a diversity of empty niches on the new continents. The molecular data suggest, however, that the breakup of land masses itself provided the opportunity for diversification.

If the molecular data are correct, fossil evidence for two dozen orders is "missing" for upward of 30 million years, and by some estimates 60 million years, of stratigraphic history. The real question asked by many paleontologists, therefore, is, "How likely is it that an increasingly good fossil record could be devoid of evidence for such diversification?" The answer appears to be, not likely. Using standard models to examine the probability of not finding a species in the fossil record, scientists estimate that, even under the most severe conditions, there is a 2% chance of not finding earlier fossils if, indeed, ordinal diversification occurred before the Cretaceous/Tertiary boundary. One possible reconciliation of these two viewpoints would be the Cretaceous splitting of a few lineages, thus giving rise to earlier ordinal diversification, followed by a rapid radiation during the Paleocene.

Whatever the timing of the ordinal diversification of mammals, by the mid-Tertiary, continental drift resulted in a relatively large number of isolated land masses, evolutionary caldrons in which mammalian diversification occurred. Peak continental isolation occurred at the Cretaceous/Tertiary boundary when landmasses, which in modern times are joined, were isolated. This included present-day continents (e.g., South America and Africa) and subcontinents (e.g., India). Even landmasses that were joined, such as Asia and Europe, exhibited greater isolation as a result of landscape features.

Island continents, in and of themselves, would have likely produced greater mammalian diversity over time. However, in addition to the isolation of continents, the Tertiary was a time of greater differentiation and seasonality of climates. Much evidence exists for this, from diversification of Tertiary flora (particularly in the late Tertiary) to developments such as the Antarctic ice cap, which formed in the Miocene. Continental drift, with landmasses moving toward the poles, combined with concomitant changes in ocean currents resulted in increasingly temperate environments. For the homeothermic mammals, this diversification of ecosystems provided a unique opportunity for radiation. In Recent times, geologically isolated continents and large islands, such as Madagascar and Australia, continue to show unique assemblages of mammals that are typified by high levels of endemism and diversification of a small number of families.

III. PHYLOGENY AND BIODIVERSITY: PATTERNS OF ORDINAL DIVERSIFICATION

Mammalian generic diversity appears to be rising exponentially throughout the Cenozoic (Fig. 3). While there is some fluctuation in the levels of diversity, the progressive increase in diversity is notable. This pattern may be driven, in part, by an increasingly heterogeneous environment. However, for the most part, the exponential rise is probably an artifact of an increasingly well represented fossil record. Mammal diversity probably peaked during the mid-Tertiary, although patterns of diversification (and simplification through extinction) vary from continent to continent. The following sections will briefly describe patterns of diversity in the extant orders of mammals.

A. Monotremes

The three extant species of the Montremata, found only in Australasia, all represent highly specialized representatives of an order that had an early branching from all other mammalian orders. Living monotremes are toothless, with the exception of young platypuses, a trait that has not made reconstruction of fossil history a simple task. Early fossil remains are known from Australia. A single fossil found in southern Argentina, dated to about 60 million years B.P., represents the only occurrence outside of Australasia.

Monotremes possess what are surely primitive mammalian characters such as egg laying and a cloacal reproductive/excretory structure. While these primitive characters reflect an early divergence, monotremes are not primitive in all their characters. For instance, all three extant species of montremes share complex and well-developed nervous systems. The most unusual sensory adaptation is the development of electroreceptors across the platypus's bill and in the tip of the echidnas, producing snout or beak. Unique among mammals, electroreception in monotremes is complemented by a complex and large brain.

Monotremes are also the only mammals that have the ability to deliver venom as a defensive measure. While venom glands are vestigial in the echidnas, in the male platypus a well-developed venom gland is found in the hind legs. More active during the mating season, the venom is assumed to be a defense against intruding males. The venom is toxic to humans and lethal to small mammals.

The sole surviving representative of the Ornithorhyncidae is the platypus, *Ornithorhyncus anatinus*, found only in eastern Australia and Tasmania. The morphology of the species is so distinctive that after early examination of skins sent to London, they were pronounced a fraud. Weighing around a kilogram and with an overall length of 45 cm, the platypus has a broad, duck-shaped bill. The animal is covered in short, dense fur and has webbed feet with claws for digging. The animal is highly aquatic and solitary outside of the breeding season. Females will expand their burrows before breeding and, after mating, will "gestate" one or two eggs for 10–12 days. After laying, the egg is incubated for a further 12 days. Lactation is protracted, approximately 4 months, but the young becomes independent soon after weaning.

The Tachyglossidae is represented by two extant monotypic genera: *Zaglossus bruijni*, the long-beaked echidna, and *Tachyglossus aculeatus*, the common echidna, or spiny anteater. *Zaglossus* is endemic to the mountains of New Guinea and feeds primarily on earthworms and arthropods, while *Tachyglossus* has a distribution across much of Australasia, found wherever its primary prey, ants and termites, are abundant. Introduction of the dingo 10,000 years ago probably has reduced the range of *Tachyglossus*, much as it has been thought to have a similar effect on the platypus. The echidnas are larger than the platypus, with *Zaglossus* weighing up to 10 kg and *Tachyglossus* reaching about half that weight. The two species have thin protruding snouts and thick coats of dark brown fur interspersed with sharp quill-like spines. Patterns of reproduction differ from those of the platypus. The egg is deposited into a pouchlike structure, where it hatches 10 days later. The young stays within the pouch for 2 months and then nurses outside the pouch for a further 3 months.

B. Marsupials

Extant marsupials are grouped into 7 orders, with 21 families, 83 genera, and 276 species (Table I). In a pattern commonly seen among mammals, some lineages contain disproportionately more families, genera, and species than others. For instance, more than half of all extant marsupial families are in the order Diprodontia, while three orders are represented by only a single family. The distributions of genera within families, and species within genera, are also highly skewed. Three orders of marsupials are found only in the New

TABLE I

Diversity of Recent Marsupials

Order	Common name(s)	Number of families	Number of genera	Number of species
Didelimorphia	American opossum	1	15	63
Paucituberculata	Rat opossum	1	3	6
Microbiotheria	Minito del monte	1	1	1
Dasyuromorphia	Thylacine, numbat, quoll, antechinus, Tasmanian devil	3	17	63
Peramelemorphia	Bandicoots	2	8	21
Notoryctemorphia	Marsupial mole	1	1	2
Diprodontia	Koala, kangaroo, wombat, wallaby, glider, leathertail, cuscus, phalanger	10	35	117
Totals		21	83	276

World, while the remaining four orders are found in Australasia. The Old World continents of Africa, Asia, and Europe have no extant marsupial representation.

1. Didelimorphia

The American possums, or didelphid marsupials, represented by a single modern family, are found throughout South America, with the Virginia opossum (*Didelphis virginiana*) found widely over eastern, central, and southern North America. A generalist capable of living on garbage, the Virginia possum has shown a remarkable range expansion in the past two centuries, expanding its range from the southeastern United States north into the central states and the southernmost fringes of eastern Canada and along the west coast up to and just over the Canadian border. In the past 50 years alone, the species has extended its range over an area of two million square kilometers. This expansion has been facilitated by the opossum's extreme potential fecundity. A female may give birth to up to 60 very altricial young, up to a quarter of which may successfully attach to one of her 15 teats.

Despite ranging in size from a small mouse to a large cat (25 g to 5 kg), most didelphids show a clear set of similarities. Arboreal habits are common in the Didelphidae as is reflected in the general morphology of family members. Hands and feet are well developed, with an opposable big toe on a hind foot that has five toes. Many species also have prehensile tails.

The dentition of most didelphids is relatively unspecialized, with 50 teeth found in most species. The family is dominated by generalists, many of which have wide geographic ranges. However, some of the more remarkable specialists are also found in the didelphids. The yapok, or water possum (*Chironectes minimus*), is the only truly aquatic species, with a fully webbed hind foot and a carnivorous diet consisting of a variety of crustaceans, fish, and amphibians.

2. Paucituberculata

Paucituberculata is also represented by a single Recent family, the Caenolestidae. All six extant species, placed in three genera of shrewlike marsupials, are found only at higher altitudes in the Andean cordillera, from Ecuador south to Chile. Capable of killing small vertebrates with their sharp teeth, the caenolestids usually subsist on invertebrates, particularly earthworms. Weighing under 20 g, the shrew opossums are nocturnal omnivores, using their long snouts and tactile whiskers, or vibrissae, to locate and kill prey. The extant paucituberculates are represented by a much more diverse fossil fauna. In the Miocene, seven genera of caenolestids were represented by a large number of species with widespread geographic distribution and a relatively large diversity of life forms, including specialized predators.

3. Microbiotheria

A single Recent species of Microbiotheria represents an order whose fossils have been found on three continents: South America, Australia, and Antarctica. With origins in the Paleocene, the microbiotherids are thought to have invaded South America while it was connected to Antarctica and Australia via a southern land bridge. The historical diversity of the group in South America is poorly known. The extant arboreal marsupial mouse inhabits the forests of southern Chile and is nocturnal.

4. Dasyuromorphia

Dasyuromorphia is known from fossil evidence in Australia in the mid-Miocene (Dasyuridae), but by this time the order already exhibited a relatively large diversity of morphotypes. Extant members of the order are represented by 3 families and 63 species, approximately 20% of all Recent marsupials. All but two of these species can be found in the family Dasyuridae.

In historical times the Thylacinidae were represented by a single species, the thylacine *Thylacinus cynocephalus*, also called the Tasmanian wolf or Tasmanian tiger. This species has been considered extinct since the last specimen died in captivity in 1936. Thought to have had a wide distribution across Australia, introduction of the feral domestic dog, or dingo, by Polynesians approximately 10,000 years ago probably led to its extinction in Australia. The cause of extinction of the relictual population in Tasmania is debated, although it was most likely the result of a number of interacting factors common to human predator extinctions: habitat loss, disease, and declining prey supplanted by sheep farmers, who hunted the thylacine because their livestock was perceived to have been threatened by predation.

The Myrmecobiidae consists of a single, monotypic genus, represented by the numbat, or banded anteater. Weighing approximately 400–500 g, the numbat resembles a small grayish-red squirrel, with a long bushy tail, a half dozen lateral white stripes circling its abdomen, and a prominent protruding snout. As its family name would suggest, however, the numbat is a termite specialist. Like other termite specialists (e.g., the bat-eared fox of Africa), the historical distribution of the species was dictated by the occurrence of its primary prey. While termites are found across southern and central

Australia, today the numbat persists in small fragments of *Eucalyptus* forest and woodlands (hardly core habitat), its last refuges from agriculture and predation by the introduced red fox (*Vulpes vulpes*).

The 61 species of the family Dasyuridae are often called the marsupial carnivores despite the fact that most members of the family are insectivores. Their dentition, while similar to that of most marsupials, is similar to the dentition of many carnivores, characterized by bladelike incisors, large, sharp canines, and upper molars modified with large, sharp cusps. Found across Australia and New Guinea, dasyurids are usually small, with most species weighing under 50 g. The family occupies a wide diversity of terrestrial niches, with the smallest species convergent on shrews (family Soricidae) found in Asia, Europe, and North America.

Some of the dasyurids, particularly species in the genus *Antechinus*, have remarkable life histories. For instance, Stuart's marsupial mouse, *Antechinus stuartii*, has a semelparous life cycle more similar to that of many insects than to that of a mammal. The species is usually arboreal but forages terrestrially for invertebrate prey. Mating is highly seasonal, with males competing aggressively for mates. Females produce an enormous litter for their size (10–12 offspring). The young forage independently at 3 months and are sexually mature by 10 months. What is most remarkable is that after mating, the males die, while females rarely survive to breed in a second breeding season. Similar patterns of reproduction are seen in other *Antechinus* species.

The two largest species of dasyurids rightfully deserve the name marsupial carnivore. The tiger quoll, *Dasyurus maculatus*, may weigh up to 7 kg (males) while the largest extant species of Australian carnivore, the Tasmanian devil (*Sarcophilus laniarius*), can reach the size of a medium-sized dog, weighing up to 9 kg. Like the thylacine, the Tasmanian devil once had a widespread distribution across Australia but is now extinct on the continent. It persists in Tasmania, perhaps because it is primarily a scavenger, not a predator, and thus is more catholic in its dietary requirements and less of a perceived threat to farmers. Both quolls and Tasmanian devils are usually solitary; however, aggregations of up to 20 Tasmanian devils have been observed around carcasses.

5. Peramelemorphia

Found from the deserts of Australia to the rain forests of New Guinea, the 21 species of bandicoots and bilbies are placed in 8 genera and 2 families, the Peramelidae and the Peroryctidae. The bandicoots are insectivorous or omnivorous and for the most part are ecological counterparts of insectivores. Most species forage by digging insects out of the ground. They have small, sharp teeth, and ecological studies show that, despite feeding specialization for insects, they are omnivorous and opportunistic in their diet choice.

Bandicoots vary in size from the rodentlike mouse bandicoot (*Microperoryctes murina*) at 15–18 cm up to the giant bandicoot, *Peroryctes broadbenti*, at 45–50 cm and up to 5.5 kg. In Recent times, the evolution of bandicoots has taken separate paths in Australia and New Guinea. Of the 11 species of peroryctids, or spiny bandicoots, only two are found outside New Guinea. One species is found on the Indonesian island of Seram, in the Molluca or Spice Island archipelago, due west of New Guinea, while another has a relictual distribution in the rain forests of Cape York in northern Australia. Little is known about the fossil history of this family as New Guinea is all but lacking in fossil record. On New Guinea, species occur at different altitudes, with the most species (six) occurring in midaltitude regions of 1000–1500 m above sea level.

Like perrisodactyls and artiodactyls, peramelid bandicoots show relatively extreme reduction in their limb morphology. Many peramelids have digitigrade elongated hind limbs, with expanded growth of the fourth digit, and lateral digits showing varying ranges of reduction. Reduction in limb digits suggests strong selection for running in the bandicoots. While the bandicoots are remarkably cursorial, some of the elaboration of the hind limb may also have evolved for improved digging and burrowing.

6. Notoryctemorphia

The only Australian marsupial adapted for a fossorial existence is the sole species of the order Notoryctemorphia, *Notoryctes typhlops*. The marsupial mole has a pale, golden yellow coat and a long, tubular aspect about 15 cm long, weighing 35–40 g. There are rudimentary, almost vestigial eyes, compressed neck vertebrae, and stubby, almost vestigial forelimbs with two well-developed claws on the third and fourth digits. Other digits have essentially disappeared. The hind limbs are flattened, with only three small claws, and are used to expel dirt from a burrow. Little is known about the behavior or ecology of the species, although it is believed to be solitary and has been observed to eat burrowing insects and their larvae.

7. Diprodontia

This order includes both the largest and smallest marsupials. The smallest member of the family Burramyidae, or pygmy possums, weighs 6–8 g (*Cercartetus lepidus*).

Known only from fossil remains until 1996, the family is now known to have five extant species placed in two genera. At the other extreme of the diprodont scale are the members of the family Macropodidae, the kangaroos and wallabies. While the smallest macropod, the rock wallaby (*Petrogale burbidgei*), weighs in at 1 kg, the great gray kangaroo may reach 90 kg and stand flat-footed at just over 2 m. Diprodontia also shows the greatest morphological variation on any extant marsupial order. While it is beyond the scope of this article to review all 117 species in the 10 families that make up the diprodonts, a review of some of the more divergent taxa is informative.

The single species in the family Phascolarctidae, the koala (*Phascolarctos cinereus*), like the kangaroo, is emblematic of the Australian continent. Highly selective in its diet, the koala subsists entirely on the leaves of a few species of smooth-barked eucalyptus trees. Unlike many marsupials, the koala is slow to reproduce, with a single young born to a female each year and the time of dependency between mother and young stretching to up 1 year. Sexual maturity is not reached until age 3 or 4.

The feathertail glider (*Acrobates pygmaeus*) and the feathertail possum (*Distoechurus pennatus*) are the only two species in the aptly named family Acrobatidae. Found only in New Guinea and with no known fossil record, both species have long stiff hairs on either side of their tails, from which they derive their name. The gliders, which weigh a mere 10–15 g, have membranes between the elbows and knees, further aiding gliding flight. The membrane is lacking in the possum, which weighs up to 50 g.

Gliding as a strategy appears to have evolved twice in the diprodonts. Five species of lesser glider, along with the striped and Leadbetter's possum, make up the family Petauridae. Found from Tasmania to New Guinea, the petaurids resemble small squirrels. Weighing from 100 to 700 g, they have long bushy tails and stripes down their backs. The striped possum has black and white dorsoventral stripes along the back, resembling a North American skunk in coloration.

The five gliders bear a remarkable resemblance to flying squirrels of the genus *Glaucomys*. A membrane that stretches between their wrists and ankles provides an almost rectangular, kitelike gliding surface that these animals use to move from tree to tree. While the striped possum is an insectivore, all other members of the family specialize on eating plant exudate (sap and gum) as well as insects. The sugar possum (*Petaurus breviceps*) is particularly specialized, concentrating its efforts on the exudate from a single species of *Eucalyptus*.

Resembling fossorial rodents, the three extant species of wombat have short, muscular limbs, stocky bodies, and broad long claws. They are large animals, weighing up to 40 kg. Prodigious diggers, they live in complex burrow systems, which caused them to be called marsupial badgers on their first discovery. Unique among marsupials, their teeth are ever-growing like those of rodents and lagomorphs. Wombats are herbivores, concentrating their diet on tussock grasses. Wombats are found only in southern Australia and Tasmania, with the northern hairy-nosed wombat (*Lasiorhinus krefftii*) found in one isolated population of 70 individuals living on 3 square kilometers in central Queensland. Wombats have a long fossil history in Australia, with two much larger species occurring throughout the Pleistocene and going extinct about 10,000 years ago, coincident with the arrival of humans and dogs.

The best known family of diprodonts, and probably the most widely recognized family of marsupials, are the 11 genera and 54 Recent species of macropods, the kangaroos and wallabies. Ranging in size from 1 to 90 kg, all species share the same basic body plan. Forelimbs are reduced and have five toes, all bearing small claws. The hind limbs are elongated, and, as in the bandicoots, on the hind leg all but the fourth digit has been reduced or lost. The large hind foot allows macropods to take a bipedal stance.

The limbs are highly adapted for ricochetal or hopping locomotion that, because of the limb structure, is highly energy efficient. Each time the kangaroo bounces down, the large elastic tendons in the foot are stretched taut like a rubber band, storing energy from the previous bounce. As the foot is released, the tendons snap back and push the foot off the ground, throwing the kangaroo into the air. On landing, the tendons are stretched taut again, completing the cycle.

In addition to the remarkable limb adaptations, macropods have a digestive system remarkably similar to that of the eutherian ruminant herbivores they ecologically replace. Both groups have evolved segmented stomachs, with a true chambered stomach in the ruminants and a divided, three-part stomach in macropods. This segmentation allows for the acquisition of foregut fermentation, in which symbiotic bacteria digest the tough outer cell walls of the plant materials that the animal eats, thus releasing nutrients from within the cell, providing a digestible form of cellulose (which is broken down by the bacteria), and nutrients from the bacteria themselves as they are absorbed when they pass through the gut.

Macropods have a wide geographic range and habits. While macropods are usually thought of as long-distance, open-plains grazers, in New Guinea tree kangaroos in the genus *Dendrolagus* have taken to an arboreal habit. While these animals still spend much of their time on the ground foraging for fallen fruits, they are agile climbers and use their large hind feet to propel them from tree to tree. Many of the smaller wallabies are also habitat specialists, such as the 10 species of rock wallabies (*Petrogale* sp.).

C. Placentals

Placentals dominate both the fossil and recent history of the Mammalia. With fewer than twice the number of orders of marsupials (18 vs 10), placentals have shown remarkable diversification at all lower levels. Recent species are grouped into 18 orders with 115 families, 1049 genera, and 4375 species (Table II). As in the marsupials, some lineages contain disproportionately more families, genera, and species than others. With a fossil record stretching back into the Cretaceous, eutherian mammals have a long evolutionary history derived from marsupials and monotremes. Found on all continents, the placentals show remarkable niche diversification, occupying a wide variety of terrestrial, aerial, and aquatic habitats and subsisting on diets ranging from the true omnivorous habits required to live on the varied detritus of human populations to a single group of grasses (pandas and bamboo) or termites.

1. Xenarthra

A group found in Recent times only in the New World, xenarthrans, are thought to have occurred in Europe, although the fossil record is, at best, fragmentary. To appreciate the diversity of the Xenarthra, the sloths, anteaters, and armadillos, take a journey back into Miocene South America. Giant sloths, commonly called ground sloths or sloth bears, were as huge as their name would indicate. Numerous families of armadillos, including the Glyptodontidae, lumbering creatures up to 3 m long, occupied the landscape. While Tertiary radiations included a variety of species exploiting numerous habitats and niches, the members of the extant families of Xenarthra are relatively or highly specialized: all four species of anteaters (Mymecophagidae) are obligate ant and termite eaters, using their long, sticky tongues to collect their food. The sloths (five species

TABLE II

Diversity of Recent Eutherian Mammals

Order	Common name(s)	Number of families	Number of genera	Number of species
Xenarthra	Sloths, armadillos, and anteaters	3	13	29
Insectivora	Tenrecs, moles, hedgehog, shrew, mole	7	66	429
Scandentia	Tree shrews	1	5	19
Dermoptera	Colugo	1	1	2
Chiroptera	Bats	18	178	928
Primates	Apes, gibbons, marmosets, lemurs, galagos, Old World and New World monkeys	13	60	236
Carnivora	Canids, felids, mongoose, hyena, seals, weasels, racoons, otters	11	129	271
Cetacea	Whales, dolphins, and porpoises	10	41	78
Sirenia	Sea cows, manatees, dugongs	2	2	5
Proboscidea	Elephants	1	2	2
Perissodactyla	Equids, tapir, rhino	3	6	17
Hyracoidea	Hyrax	1	3	6
Tubulidentata	Aadvark	1	1	1
Artiodactyla	Pigs, peccaries, hippos, deer, antelopes, bovids	10	81	220
Pholidota	Pangolins	1	1	7
Rodentia	Rodents	29	443	2024
Lagomorpha	Rabbits and pikas	2	13	81
Macroscelidea	Elephant shrew	1	4	15
		115	1049	4370

in two families) have acquired an odd but apparently effective strategy, shifting along tree branches upside down, their spines hanging in a catenary curve. Sloths are the only green mammals; however, their color is not intrinsic to their fur but is derived from algae and cyanobacteria that grow in grooves in their hair. Sloths come down to the ground only to defecate. Armadillos (20 species in the family Dasypodidae) are distinguished by their jointed armor.

2. Insectivora

Modern insectivores show a wide range of adaptations and are found on every continent except Australia. Taxonomy of the insectivores has been problematic, not only because the order contains some of the most primitive eutherian mammals, as well those with highly derived characters, but also because the order was used as the equivalent of a taxonomic wastebasket in which problematic families were thrown. The 6 families and 66 genera now classified as insectivores form a monophyletic (single-root clade) with ancient origins in the late Cretaceous or Paleocene.

Two families of insectivores, Solenodontidae and Nesophontidae, were found in modern times only in the Caribbean. All of the Nesophontidae (also called West Indian shrews) are extinct. The two remaining *Solenodon* species are found in Haiti and, tenuously, in Cuba. Expansion of agriculture and introduction of rats, mongoose, and companion animals (dogs and cats) have led to decline of these species. *Solenodon* are shrewlike, have a distinctive, highly flexible snout and large hind feet, but are distinguished by an unusually large body size, weighing up to a kilogram. Fossil solenodons are known from North America from Oligocene deposits, approximately 30 million years B.P.

Tenrecs (family Tenrecidae) show a disjunct distribution, found on islands of the western Indian Ocean (Madagascar, Comorros) and central Africa, and have probably been linked to Africa for all of their evolution. The family can best be described as diverse. While only ranging in body size from a shrew to a small cat (maximum weight of a kilogram), tenrecs have undergone a remarkable radiation on Madagascar, expanding into vacant niches usually occupied by animals as diverse as otters, shrews, hedgehogs, and moles. Many species are omnivorous.

The family Chrysochloridae, the golden moles, is restricted to southern Africa. The 18 species in 7 genera are all fossorial and closely resemble both true moles (also insectivores, family Talpidae) and the marsupial mole, Notoryctidae. While the entire evolutionary history of the chrysochlorids is in eastern and southern Africa, true moles, or talpids, are found throughout the Northern Hemisphere and have long evolutionary histories in both Europe and North America. The talpids are more diverse than the chrysochlorids, with a range of morphotypes including the star-nosed mole (*Condylura cristata*), which has nearly two dozen small, fleshy fingers arranged in a starburst pattern on its nose. This represents the most bizarre elaboration of a general phenomenon found in the true moles, which have thousands of receptors arranged in a structure called the Eimer's organ on the snout. Cortical development of the brain is linked with the Eimer's organ. The Eimer's organ assists moles in navigating and foraging in their tunnel networks.

3. Scandentia

The 5 genera and 19 species of tree shrews (family Tupaiidae) are the only extant members of the Scandentia. The group is dispersed across the rain forests of southern Asia and Southeast Asia and has a relatively long fossil history (first found in the Eocene of Asia). Resembling a small squirrel, weighing 45–50 g, tree shrews have an elongated snout and a long, bushy tail. Like true shrews, many species are known to consume insects; however, the few scientific studies that have been conducted on the family have found fruit is often the dominant component of the diet. Parental care is unusual, with the female giving birth to a small number of young (1–3) and hiding them in a nest. In a pattern more common in antelopes, the female visits the young to nurse and then leaves them lying in. Maturation is rapid, with independence from the mother at about a month and full sexual maturity at 4 months. Tree shrews are enigmatic in their phylogenetic affiliations. They have variously been assigned to the primates, flying lemurs (Dermoptera), and rabbits (Lagomorpha), with both molecular and morphological analysis producing contradictory results.

4. Dermoptera

Gliding, as opposed to true flying, is a mode of locomotion that has evolved independently in a number of mammalian orders. The two living species of Dermoptera (family Cynocephalidae), also called colugos or flying lemurs, are a relictual group of gliders found today only in the forests of Southeast Asia. Not closely related to lemurs, the dermopterans first appear in the fossil record in the Eocene of Thailand. Throughout their evolutionary history, only two families of colugos have been described, one North American and the other Asian. Affiliations of the order are unclear, with some paleontologists placing them as relatively close relatives to bats: they are not related to lemurs despite their common name. The gliding membrane, or patagium, stretching over the animal's entire body, connecting along both sides of the tail to the legs, from legs to arms and from the arms to the neck, makes the animal look like a kite. Despite their relatively large size, reaching a maximum size just under 2 kg, the colugos are spectacular gliders, moving over 100 m from tree to tree in the forest to reach the trees on which they feed. The colugo diet is herbivorous but may include leaves, fruits, and flowers.

5. Chiroptera

Bats represent the second most diverse order of mammals. Two major suborders of bat have been distinguished, the Megachiroptera, also called fruit bats or flying foxes (166 species in 42 genera), and the Microchiroptera (762 species in 136 genera). Bats come as small as 1.5 g with a wingspan of 15 cm (Craseonycteridae, Kitti's hog-nosed bat) and as large as the flying foxes in the genus *Teropus* (Teropidae), which may weigh up to 1.5 kg and have wingspans of 2 m.

The two suborders have long, independent evolutionary histories, although the very nature of their morphology, and in particular their fine, light bone structure, makes preservation unusual. In contrast to modern diversity, sufficient fossil material has been found to describe just over 30 genera. The lack of fossil evidence has provoked repeated controversy and called into question whether bats are monophyletic or whether the fruit bats are actually more closely related to another group (usually primates are postulated because of shared evolution with megachiropterans of details of their neural pathways for vision). If this were the case, the remarkable flight structures would have to have evolved twice independently. Recent molecular data suggest, relatively unambiguously, that the two bat groups are monophyletic.

Bats are the only mammals capable of true flight (as opposed to gliding). The wing is formed by a thin membrane that stretches across the arms, elongated fingers, and along the body, forming a diaphanous umbrella-like structure, or patagium. Like all flying animals, bats need wings that are at once light but strong. The wings are greatly reduced in weight with muscles pulled in close to the body and bones reduced in size and volume. Torsional stress is reduced by simplification of joints. The remarkable flying abilities of bats ensure a global distribution, with bats often the only mammal naturally occurring on remote islands.

While birds monopolize diurnal aerial niches, bats are nocturnal specialists. In some areas where birds and bats do not co-occur, bats may become more diurnal, providing thin evidence of competitive exclusion. Flying at night, visual acuity is of relatively little value; nonetheless all but one species of Megachiroptera rely on vision to navigate. In contrast, microchiropteran bats have evolved navigational tools that are independent of sight. Using their larynxes, microchiropterans produce extremely high frequency sound that they emit through their nose or mouth. The sound produced by bats is referred to as ultrasonic because it is above the range that humans can hear. The ultrasound bounces off both potential prey and obstacles and is received back at the large, elaborated ears of the bat. Because the bat navigates using the sound that bounces back, this form of navigation is called echolocation.

Despite their high specific diversity, patterns of reproduction (in the few species that have been studied) all are remarkably similar. Females carry their young with them while in flight, and this constraint limits bats to producing one, or occasionally two, offspring. Paternal care of offspring is rare; hence social systems tend to be promiscuous, with few lasting bonds between males and females. Bats tend to be highly gregarious, hanging upside down in roosts containing hundreds, thousands, and in rare cases hundreds of thousands of individuals. Caves, cliffs, and the eves of large houses all provide appropriate roosting areas.

Bats are usually insectivorous, gleaning their prey while airborne, but the diversity of diets is enormous. Diet specialists cover a range of vertebrates and include species that are adapted to fish, much like a small eagle, frog eaters, blood-eating specialists (the vampire bats), nectar feeders, and two groups of specialized fruit eaters. Two families of bat dominate fruit eating: in the Old World, Pteropidae (flying foxes), and in the New World Phyllostomidae (spear-nosed bats). In the forests in which they live, fruit-eating bats are keystone species, dispersing many of the largest forest seeds and fruits. Of course, a frugivorous diet and the large aggregations in which fruit-eating bats can occur make some bats extremely unpopular with fruit farmers around the tropics.

6. Primates

In Recent times, primates have been found on every continent and in every habitat on earth. If one excludes the most abundant and successful primate, humans (*Homo sapiens*), the range of the order is much reduced, with representatives found in the tropics and subtropics of Africa, Asia, and Latin America. One macaque species (Barbary macaques, *Macaca sylvanus*) is found in Europe on the island of Gibralter across narrow straits from Africa. Another macaque species, the Japanese macaque, *M. fuscata,* survives in a temperate environment by relying on hot springs during the winter.

With the first fossil representation in the late Cretaceous of North America, primates are an ancient order. The modern order is represented by 13 families, including the apes, gibbons, marmosets, lemurs, and galagos. Just over half of the familial diversity is found in the suborder Strepsirhini, or prosimians, which includes the five families of lemurs, all of which are endemic to the island of Madagascar, the Loridae (pottos and lorises), and the Galagonidae (galagos), small nocturnal monkeys of Africa often called bushbabies. Some include the Tarsiriidae as well.

The 30 species of lemurs show a remarkable range in size from 50 g (lesser mouse lemur, *Microcebus murinus*) to over 10 kg (Indri, *Indri indri*) and a diversity of diet from generalized herbivores, insectivores, a larvae specialist (the aye-aye, the only extant species in the family Daubentonidae), and three species of bamboo specialist (*Hapalemur* species). Many members of this suborder are under threat, victims of the rapid and seemingly irreversible deforestation of Madagascar.

While 11 species in 4 genera of galago are recognized, the specific status of the family Galagonidae is in question. Recent, and as yet unpublished, genetic data indicate that within a single species of galago there may be several cryptic species, virtually indistinguishable morphologically from one another.

The remaining primates are placed in the suborder Haplorhini, which includes the five families of New and Old World anthropoid primates and the tarsiers. Found across the forests of Southeast Asia, the tarsiers bear a strong physical resemblance to galagos and for many years were grouped with them in the prosimians. Recent genetic evidence, and a reconsideration of morphological characters, place the tarsiers firmly as distant, but distinct, relatives of the anthropoid primates.

The extant South American primates fall into two families, the Callitrichidae (tamarins and marmosets, 26 species) and the Cebidae (58 species, including capuchin, squirrel, howler, spider, and night monkeys and the sakis). Marmosets and tamarins are unusual in that all but one species regularly produces twins. They have a monogamous or polyandrous mating system (one female, several males), and a diet that is highly omnivorous and may include fruit, seeds, insects, and small vertebrates. Some species, including the diminutive pygmy marmoset (150 g), are exudate specialists, eating the gum of trees.

The Cebidae have been described as monkeys with five legs due to the extensive use they make of their prehensile tails, an adaptation not found in Old World monkeys. Body size ranges from just over 500 g for squirrel monkeys (*Saimiri sciureus*) up to 12 kg for the wooly spider monkey (*Brachyteles arachnoides*). The diet is predominantly fruits and other vegetation and competition for resources is common among cebids, with larger species dominating smaller ones. As a result, many smaller species have evolved diet or behavioral specialization, including nocturnal foraging (night monkey, *Aotus trivirgatus*), the ability to eat green fruit (titi monkeys, *Callicebus* sp.), or habitat specialization (swamp living in uakaris, *Cacajao* sp.).

Old World monkeys in the family Cercopithecidae dominate the primates numerically, with 81 Recent species in 18 genera. Found across Asia and Africa, cercopithid monkeys include animals with divergent life history strategies, sizes, and diets. The guenons, macaques, and baboons (subfamily Cercopithecinae) range in size from 1 to 50 kg and are usually found in large, extended matrilineal groups. Many species are tree-living, but the baboons and some macaques have adopted a terrestrial or semiterrestrial existence. Primarily fruit eaters, most species are omnivorous and will supplement their diet with practically anything that they can capture and eat.

The columbine monkeys, langurs, and leaf monkeys are placed in the subfamily Colobinae. A diverse group of animals, this group tends to be more arboreal and slighter than the cercopithicines. While the evolutionary history of this group is mainly African, modern representation is dominated by Asian forms (4 genera, 28 species). Diet is reflected in an unusual stomach structure, with a large, saclike upper stomach that, like the rumen of ungulates, allows for fermentation of the coarse leaves that constitute much of this group's diet.

The 11 species of gibbons (family Hylobatidae, genus *Hylobates*) are commonly called the lesser apes and are often grouped together with the true apes, or Hominidae. Found throughout tropical Asia, gibbons are renowned for their brachiating (or swinging) locomotion and their echoing, eerie loud vocalizations. The family is primarily monogamous, and its diet is primarily fruit, supplemented with young leaves and flowers.

The best studied group of primates is the family Hominidae. Formerly broken into two families, Pongidae (orangutans, gorillas, chimps, and bonobos) and Hominidae, recent data, particularly molecular genetic analyses, have made the isolation of human beings and their ancestors in their own family impossible. While largely vegetarian (gorillas are almost exclusively folivores), some chimps have been known to actively hunt other monkeys and to kill conspecifics. While most species are extremely social, orangutans tend to be solitary except when mothers are in association with young or during the mating season. Slow breeders and habitat specialists, great apes are declining in number and distribution as a result of the rampant expansion of their con-familial *Homo sapiens*.

7. Carnivora

A diverse group of animals, the Carnivora share a key diagnostic feature. The carnivore dentition is typified by the evolution of a single pair of slicing teeth, the carnassials. Formed from the upper fourth premolar and the lower first molar, the carnassials are the key to understanding the diversity and plasticity of Carnivora. By isolating the shearing function to a single set of teeth, true carnivores have retained flexibility in their diet while allowing for specialization of shearing of meat. Molars retain their grinding function, while premolars in front of the carnassials are retained for grasping, crushing, or puncturing food. The canines are often elaborated into spikes of one size or another, allowing a sharp stabbing of prey. While carnassials were critical to carnivore evolution, many extant members of the order have reverted to more generalized dentition and lost the carnassials (e.g., pandas grinding bamboo).

Modern carnivores are divided into two major groups, the Feliformia, which includes hyenas, felids, mongooses, and viverrids, and the Caniformia, a diverse group that includes seals, sea lions, walruses, canids, bears, procyonids, and the mustelids.

Pinnipeds are closely related to other carnivores, but there has been a running debate as to the relationship of the three families: the eared seals (Otaridae, 14 species, 7 genera of eared seals and sea lions), the earless seals (Phocidae, 19 species and 10 genera of seals), and the walruses (Odobenidae, 1 species). For many years, these taxa were considered monophyletic, but a recent suggestion was made they are diphyletic, with walruses and sea lions branching off from a common ancestor with bears and phocids (seals) being more closely related to the mustelids. More recent molecular and morphological work confirms the monophyly of the pinnipeds.

The Canidae include 34 species of foxes, wolves, jackals, and dogs. Ranging in size from the fennec fox (*Vulpes zerda*, just under a kilogram) to the wolf (*Canis lupus*, up to 80 kg), canids show a wide range of diet and social organization. The family is best known for its pack-living, social animals (wolves, the Asian wild dog or *Cuon alpinus*, the African wild dog *Lycaon pictus*, and the South American bush dog *Speothos venaticus*), species that raise a single litter of pups communally and that are capable of hunting animals up to 10 times their body weight through coordinated group hunting. Most species, however, live in small family groups and subsist on rodents and small vertebrates. One species, the African social bat-eared fox (*Otocyon megaloitis*), specializes on termites. Another species, the stilt-legged South American maned wolf (*Chrysocyon brachyurus*), is a solitary omnivore.

Mustelids, with 65 Recent species in 25 genera, are among the most diverse of the modern carnivores. A family that includes weasels, badgers, otters, and more often than not the skunks, Mustelidae are found from above the Arctic Circle to the tropical rain forests and on every continent except Australia. The group includes efficient terrestrial hunters of small mammals (mustelids, badgers) and species that are oceanic and specialize on bivalves for their diet (the sea otter, *Enhydra lutris*). Otters have become the dominant aquatic carnivore in many lakes and rivers across much of the globe. The giant otter, *Pteronura brasiliensis*, of the Amazon basin is the largest otter, weighing 30 kg, and is highly social, "herding" fish to improve success of prey capture.

Absent from South America, the nine species of badger are mostly nocturnal omnivores of stocky build and with strong, sharp claws. The Eurasian badger (*Meles meles*), a highly social animal that lives in extended burrows or setts, is an earthworm specialist. Weasels and polecats are found from northern Greenland to the tip of Africa. While the European weasel (*Mustela nivalis*) can be as small as 50 g, most weasels follow a remarkably common body form. Predominantly terrestrial and weighing 1–2 kg, with long tubular bodies and short legs, they get their speed from the compression–extension of the body; sharp teeth deliver a killing blow to a number of small vertebrates, such as frogs, rodents, rabbits, or birds. Some species once trapped for their fur are now raised in large commercial farms—e.g., mink (*Mustela* sp.).

The best known felid is the largest, the lion (*Panthera leo*). In many ways, the lion is an anomalous cat. While most of the 36 species of cat are solitary hunters, the lion is a highly social, cooperative hunter. Even congeneric large cats such as the leopard (*P. pardus*, Asia and Africa), the tiger (*P. tigris*, Asia), the snow leopard (*P. uncia*, central Asia), and the jaguar (*P. onca*, South and Central America) are more similar in their social structure and hunting patterns (solitary, territorial, with male territories overlapping those of females) to the smaller cats in the genera *Felis*. All cats are highly carnivorous, but diet ranges from the largest bovid species to frogs, fish, and mollusks. While most are nocturnal stealth predators, the cheetah (*Acinonyx jubatus*) is a true diurnal pursuit predator, capturing prey with bursts of speed of up to 90 km/h.

Viverrids (civets and genets) and mongooses (Herpestidae) are often classified as a single family, although the most recent phylogenies separate them. Both families are exclusively Old World. Viverrids show a wide range of habits, many are truly arboreal, but there is a semiaquatic form (otter civet, *Cynogale bennettii*) and a terrestrial species from Madagascar that resembles a medium-sized cat (*Cryptoprocta ferox*). Mongooses are either solitary, nocturnal predators that resemble polecats or highly social, diurnal insectivores.

8. Cetacea

The recent discovery of Eocene fossil whales confirms an early branching of the Cetacea from the primitive ungulates, closely related to modern artiodactyls. The 10 families of whales, dolphins, and porpoises show a wide diversity of body size from the finless porpoise, *Neophocoena phocoenoides*, at 30 kg, to the world's largest mammal, the blue whale, *Balaenoptera musculus*, at 150 metric tons. There is an equally large diversity of diet from tiny shrimp or krill to seals, fish, and squid—practically anything that swims. Social structure is variable and ranges from solitary individuals to the communal pods of the Orca, *Orcinus orca*, or killer whale, whose social system more closely resembles that of wolves than other whales. Cetaceans share many adaptations critical to true aquatic living. The skulls are "telescoped," with the premaxillary and maxillary bones forming the roof of the skulls and the occipital bones forming the back of the skull. In all whales the tail fin is horizontal.

Whales are divided into two suborders, the Mysticeti, or baleen whales, with 11 Recent species in 4 families, and the Odontoceti, or toothed whales, which includes 67 species in 6 families. Baleen whales, which include the aforementioned blue whale, are filter feeders. Incongruously, these leviathans live on tiny prey, from krill and small planktonic copepods, often less than 1 cm in length, to small, schooling fish. Migratory patterns across entire oceans from summer to winter feeding grounds are not uncommon, with many baleen whales being truly cosmopolitan species. Baleen whale populations, severely decreased by aggressive hunting during the nineteenth and twentieth centuries, are still in the process of recovering.

The toothed whales in the family Odontoceti are found in all oceans and all seas. Some of the smaller members of the family are also found in rivers and lakes on all continents except Australia. Like bats, the toothed whales all use echolocation, allowing them to hunt and live in turbid waters where visibility is low. While echolocation "pings" or clicks are used for locating prey, longer, more continuous tones are also important for communication among conspecifics. The toothed whales include three species of sperm whale, including the giant sperm whale (*Physter catodon*, family Phsteridae) made famous by Herman Melville in *Moby Dick*; six species of Phocoenidae, or porpoises, found throughout the Northern Hemisphere and in near-shore environments of South America; nineteen species of the poorly known beaked whales (Ziphiidae), some of whose species have never been seen alive; and two species of highly gregarious Monodontidae, limited to the Arctic oceans, the beluga whale (*Delphinapterus leucas*), or white whale, and the narwhal (*Monodon monoceros*). The narwhal has a long, spirally grooved single upper tooth, used by males in intrasexual competition, which may provide the origin of the unicorn myth.

9. Sirenia

The four species of dugong and manatee (one species *Dugong dugong* and three species of manatees in the genus *Trichechus*) are the sole modern representatives of the Sirenia. An anomalous, nearly hairless group of mammals weighing up to 1600 kg, the common name for these animals, "sea cows," is remarkably accurate. The only large, fully aquatic mammal that lives on grasses, the sirenids have extremely long intestines. Like perissodactyls, cellulose fermentation takes place in the hindgut, enabling the sea cows to process large volumes of relatively coarse seagrass vegetation. Reproduction is both delayed and slow, with long interbirth intervals and the production of a single offspring that remains with its mother for up to 2 years.

Sirenids have a near-global distribution in tropical and subtropical waters. Manatees are found up the Amazon and in major river systems of West Africa. They suffer little competition for their main resource, aquatic grasses. Yet, despite this, all species are threatened with extinction. These large mammals are docile, represent a large package of meat to human hunters, and are supposed to taste good. This, combined with slow reproduction, has led to overharvesting throughout their range. Pollution, dam building, and injury from the propellers of motorboats all amplify the problems of overhunting.

10. Proboscidea

Elephants were widely spread across North America, Asia, and Africa until recent times. Throughout the Tertiary, proboscideans showed a relatively wide range of adaptations and a diversity of species. Today, only two species remain, the African elephant (*Loxodonta africana*) and the Asian elephant (*Elephas maximus*). These hulking creatures are the largest terrestrial mammals, weighing up to 6 metric tons. While the Asian elephant is now limited to the forest of Asia, this probably reflects a historical shift from open grassland and woodland habitat which has now almost universally been converted to human agricultural use. In Africa, elephants are found throughout the forests, woodlands, and savannas, but at highest densities in savanna woodlands. Two subspecies have been defined, the forest and savanna elephant, but many ecologists believe this reflects a more recent fragmentation of populations, with a former, clinal structure of genetic variation now lost.

In both Africa and Asia, elephants are ecological keystone species, one of the few species, other than humans, having a direct impact on the form and structure of the environment in which it lives. Elephants keep savanna ecosystems from reverting to woodland when they knock down trees. "Catastrophic" destruction of trees may occur in droughts but is clearly part of a long-term natural cycle. In Asia and Africa, the fruits and seeds eaten by elephants are transported and then defecated, seeding the plains and the forests and depositing the seeds in their own fertilizer.

Elephants, fossil and modern, have a highly specialized dentition with six cheek teeth on either side of the jaw. At any one time, one to two teeth are exposed in the jaw. As these teeth wear down, they are replaced by a tooth irrupting from behind the current tooth. This pattern of sequential replacement of grinding teeth is unique in mammals. Tusks are formed from elongated upper incisors.

Highly social, female elephants live in extended, matriarchal groups. Adolescence is prolonged, with males remaining in the herd until they are 10–15 years old and females not breeding for the first time until they are 15 years old. Males are solitary, and breeding priority is determined by an interaction of sheer size (which increases indeterminantly with age) and an endocrinological sexual state called musth. Found in both elephant species, musth is associated with heightened testosterone levels and makes males extremely aggressive. The length of musth increases with age, amplifying the reproductive success of older, larger males.

Elephants have been hunted for their ivory, and in several populations in Uganda and southern Africa, tusklessness in females and reduced tusk size in males have occurred as a result of repeated selection through hunting. While poaching for ivory is an immediate threat to elephants, the real threat is loss of habitat and conflict between man and elephants.

11. Perissodactyla

If you are interested in horses, tapirs, and rhinos, the mid-Tertiary was the time to be alive. While modern perrisodactyls are represented by only eight species of equid in one genus (*Equus*), four species of tapir in one genus (*Tapirus*), and five species of rhinoceros in four genera, this represents but the faintest hint of the diversity of this order in the past. Of 14 Cenozoic families, only three families and five genera are represented in the modern mammalian fauna.

Modern equids all look pretty much the same: highly cursorial with a single hoof and consistent compact body, long neck, and highly hypsodont dentition in a large jaw with a distinct gap or diastema. Modern equids are only found in Africa and Asia, although there is evidence of the historical extinction of the tarpan, or European wild horse. Equids went extinct in North America in the Pleistocene, but asses and horses have been reintroduced in many parts of North America, the center of evolution of the Equidae. In contrast, the Pliocene of North America was populated with a few low-crowned, three-toed equids and three-toed horses, hypsodont species of various sizes, from the small, gracile *Nannippus* species to the larger *Hipparion* species. The first one-toed horse, *Pliohippus*, lived side by side with the hippo-like rhino *Teleoceras*, while in Asia the elephant-sized rhinoceros, *Sinotherium*, coexisted with the hippo-like rhino *Chilotherium*.

Only two forms of social organization are seen in modern equids. Those species that live in relatively arid environments, the asses and the Grevy's zebra (*Equus grevyi*), exhibit flexible associations among females and male territoriality. Females move widely in search of food and water, while males tend to aggregate around water holes, the limiting resource. Mating is promiscuous. In contrast, horses and the two remaining zebra species, the mountain zebra, *E. zebra*, and the plains zebra, *E. burchelli*, form harems of unrelated females and a single male, a social structure also seen in gorillas.

All five species of rhinoceros are threatened with extinction, the victim of the value of their horn. Two uses of rhino horn have been identified: shavings of the horn are used in traditional Asian medicine to bring down fever, while in Yemen, the highest value has been placed on dagger handles made from rhino horn. Income from a booming oil economy in the 1970s led to increased demand from Yemen, while economic growth in Asia in the 1980s and 1990s made rhino horn, always expensive, accessible to many more people, thus increasing demand. The black rhino (*Diceros bicornis*) was found widely across eastern and southern Africa until the late 1970s, now all but extinct except in reserves and conservancies. The white rhino (*Ceratotherium simum*), almost extinct in South Africa at the turn of the century, has recovered to where populations number 10,000. Outside of South Africa, however, the northern subspecies is all but extinct, and introduced populations in Zimbabwe were decimated in the early

1990s. The Sumatran rhino (*Dicerorhinus sumatrensis*) and the Javan rhino (*Rhinoceros sondaicus*), formerly found across Southeast Asia, are also nearly extinct. Of Asian species, only the Indian rhinoceros (*Rhinoceros unicornis*) exists in any numbers, and only as the result of major conservation efforts.

12. Hyracoidea

Modern hyraxes are classified in a single family with three genera and six species. Limited to Africa, fossil hyraxes are also found in the Middle East and Europe. While modern forms are all more or less rabbit-sized (2–5 kg) and resemble a very large guinea pig, some extinct forms weighed up to 50–70 kg. Tree hyraxes (*Dendrohyrax* sp.) are entirely arboreal, while both bush hyraxes (*Heterohyrax*) and rock hyraxes (*Procavia*) live on talus slopes, rock outcrops, and cliffs. Affiliations with other animals are uncertain, although it appears that hyraxes represent an early branching from ungulates, perhaps sharing an early evolutionary history with elephants.

13. Tubulidentata

Found across most of sub-Saharan Africa, the nocturnal and solitary aardvark (*Orycteropus afer*) is the last survivor of the Tubulidentata. Never a very diverse order, the majority of the aardvark's evolutionary history is in Africa and Europe, although aardvarks are known from Asia. Aardvarks weigh about 50 kg and resemble pigs with extremely long snouts, large ears, and long fleshy tails. Like most animals specialized for eating ants or termites, aardvarks have long, sticky tongues that are used to collect their prey. Powerful arms, with large claws, are used to excavate burrows and search out food.

14. Artiodactyla

The decline of the perissodactyls coincides with the rise of the artiodactyls. Starting in the Miocene, artiodactyls showed a remarkable diversification of families, many of which are still extant. Of 36 families found throughout the Cenozoic, 10 families and 81 genera are still found, with a near-global distribution in North and South America, Europe, Asia, and Africa. Found in all habitats, from the far north of Greenland to the southern tip of Tierra del Fuego, artiodactyls are clearly the most successful of the extant ungulates.

Diagnostic characters of artiodactyls are found in both hard and soft tissues. All members of the order have two- or four-toed hooves, with a plane of symmetry passing through the third and fourth toes. In four families (Suidae, Hippopotamidae, Tragulidae, and Tayassuidae), four fully functional digits are present, although peccaries have reduced digits on the hind limbs. All limbs have springing ligaments that capture energy when the leg flexes and return the energy to the foot as the limb pushes off, thus increasing efficiency of locomotion.

The suborder Ruminantia is represented by six families: Tragulidae, mouse deer; Giraffidae, giraffes and okapis; Moschidae, musk deer; Cervidae, deer; Antilocapridae, pronghorns; and Bovidae, cattle, sheep, goats, and antelopes. The suborder is typified by ruminant digestion, in which the rumen, an expanded first segment of a four-chambered stomach, provides an environment for the bacterial digestion of cellulose. This allows the animal to make use of highly siliceous grasses and vegetation. The cell contents, liberated by digestion in the rumen, and dead bacteria are digested farther along in the intestines. Ruminants also typically lack upper incisors, with lower incisors used to scrape grasses against the maxilla.

The tragulids, or chevrotains, are small, primitive ungulates, represented by four species in two genera, found in the forests of Africa, southern Asia, and Southeast Asia. Weighing 3–12 kg, the family lacks horns although the upper canines are unusually large and grow continually. Canines are used in intrasexual competition by males and may be used in antipredator defense by both sexes.

Cervids, or deer, are found across the Northern Hemisphere, in most of South America, and in the northern fringes of Africa. The family is absent from sub-Saharan Africa, perhaps excluded by the remarkable radiation of the antelope fauna in that region. Antlers are a defining character and are found in all but 2 of the 43 extant species. With the exception of caribou, only males carry antlers. Unlike horns, antlers are grown and shed annually. Antlers are found in a variety of forms, from spike or stubs to the elaborate branched and palmate forms seen in the moose (*Alces alces*). Deer are found in nearly every biome, from the snowfields of the Arctic, where they subsist on lichen, to the tropical forests of South America and Asia, where fruit is a common dietary staple.

The four species of musk deer in the Moschidae are found in central Asia, the Himalayan plateau, and east through parts of China and Vietnam. A primitive deer, the family was originally subsumed under the Cervidae. Lacking antlers, they have sharp, swordlike upper canines, similar to those found in the chevrotains. Musk is produced by a gland in its abdomen. This waxy substance is used as the base for many expensive perfumes and demand has led to near extirpation of all members of the family across much of its range. While captive breeding of musk deer by the Chinese might reduce demand on wild populations, prospects are not good for these animals.

Antilocaprids are only found, and have only been found, in North America. Although the family has a long and relatively diverse fossil history, only one species (*Antilocapra americana*) survives. The horns of the modern species are relatively simple when compared to the much more elaborated fossil forms. Unlike giraffids and all bovids, *Antilocapra* sheds its horn sheath annually, much like a deer losing its antlers.

The Giraffidae were more speciose during the mid-Tertiary and are represented in modern times by only two species, the giraffe, *Giraffa camelopardalis*, and the okapi, *Okapia johnstoni*. The family, exclusively Old World, has a long fossil history in Europe, where the modern genus *Giraffa* is first found in the Pliocene. Having evolved in Europe, giraffes dispersed to Africa. While modern forms have two, short stubby horns, fossil giraffes such as the genus *Giraffokeryx* had four long horns. Camel-like forms, resembling more closely the modern okapi, were also found in Europe.

By far, the most diverse of the ruminants is the family Bovidae. With 45 genera and 137 species, the family includes the greatest radiation of modern herbivores. Bovids are more widely dispersed than any other ungulate family, with a near-global distribution, absent only from Australia and the oceanic islands of Asia. But it is in Africa that the family has its greatest diversity. Members of all subfamilies are found in Africa, including 17 species of duiker (Cephalophinae, exclusively found in Africa); 35 of 40 species of antelope (Antelopinae), all 7 species of hartebeest and wildebeest (Alcelphinae), and the 1 species of impala (Aepycerotinae). The only subfamilies of the Bovidae that are relatively poorly represented in Africa are the Bovinae (11 of 24 species: African buffalo, eland, and *Tragelaphus* sp.) and the Caprinae (4 species in North Africa, none south of the Sahara).

The suids, peccaries, and hippos make up the last suborder of Artiodactyls, the Suiformes. Despite their global distribution, wild pigs and boars (family Suidae) are an Old World family, with feral populations introduced into North and South America, Australia, Tasmania, and New Guinea. They range in size from the pygmy hog (*Sus salvanicus*) of the Himalayan foothills (9 kg) to the giant forest hog (*Hylochoerus meinertzhageni*) of central and East Africa (275 kg). With 16 species in 5 genera, there are relatively few modern pig species, all of which closely resemble the domestic hog. Omnivores, pigs usually are forest dwellers, although the warthog (*Phacochoerus aethiopicus*) is found throughout the savanna woodlands and grasslands of Africa. Lacking horns, the upper canines of most pigs are ever-growing and form large, slashing tusks. This adaptation has gone to its extreme in the island endemic, the babirusa (*Babyrousa babyrussa*) of Sulawesi, Indonesia, whose tusks curl up over the head.

Closely related to the pigs are the three monotypic genera that represent the extant peccaries (family Tayassuidae). Thought to be derived from Old World pigs and with a fossil record in the Old World and New World, the peccaries are South America's pig equivalent. Smaller than pigs (20–40 kg), peccaries are critical frugivores and seed eaters thought to be important for seed dispersal in South America's tropical forests.

The third family in the suborder Suiformes includes the two species of extant hippos, the hippopotamus (*Hippopotamus amphibius*), found in Africa from the Nile basin south throughout West, East, central, and southern Africa, and the pygmy hippo (*Choeropsis liberiensis*), found only in a small section of West Africa. Recent fossil hippos are known from Madagascar, and in the Miocene, the family was spread throughout Africa and Asia. The hippo is a nocturnal grazer, leaving rivers and lakes at night to forage along their banks.

The first fossil remains of the Camelidae, the family that includes camels and llamas, are in the Old World, although by the Eocene there are camelids in the fossil record of the Americas. Former diversity was much greater than that of the present, where three species in the family *Llama* and one *Vicugna* species are found in South America, and two *Camelus* are found in the Old World: the dromedary (*C. dromedarius*, southwest Asia and North Africa) and the Bactrian, or two-humped, camel (*C. bactrianus*, Mongolian steppe). Camels are highly adapted for desert life. The feet are splayed to prevent the large (650 kg) animals from sinking into the sand, a reversal of the usual ungulate simplification of foot structure that is evident in the South American camelids and most fossil forms. Water conservation is achieved through well-insulated bodies, a high tolerance for dehydration, dry feces and concentrated urine, and metabolic conversion of fat (stored in the hump).

15. Pholidota

Pangolins, also called scaly anteaters, are not very diverse—seven species in a single genus, *Manis*—yet they have a wide geographical range, occurring in a variety of habitats across southern Africa and in much of Southeast Asia. The order is also found in the Tertiary fossil record of all continents except Australia and Antarctica. Resembling a cross between an armadillo and an anteater, pangolins are large (5–35 kg) and both ecologically and culturally important. The scales, which cover the body completely, may constitute up to half the animal's body weight and are prized for their medicinal value across Asia, although no evidence exists that the scales, which are composed, like rhino horn, of agglutinated hair, have any therapeutic value. Pangolins have the distinction of being the only extant mammals with no teeth.

16. Rodentia

The diversity of the order Rodentia deserves a volume of this encyclopedia of its own. With nearly half of all mammalian species (2024), over 400 genera, and 29 families, rodents are the most successful group of mammals living today. One family alone, the Muridae, includes two-thirds of the living species (hence, one-third of all mammals) and is subdivided into 17 subfamilies. The order includes rats, mice, squirrels, guinea pigs, beavers, kangaroo rats, dormice, jerboas or jumping mice, hamsters, mole rats, porcupines, chinchillas, agoutis, and nutria. Rodents are used by humans for food (e.g., guinea pigs), fur (nutria, beaver, chinchilla), and pets. Most rodents weigh about 100 g and have relatively similar body plans: long nose, large eyes, and long tails. Size varies four orders of magnitude, however, with the smallest member of the order weighing about 5 g (pygmy mouse, *Baiomys* species of Central America) while the largest, the capybara (*Hydrocherus hydrochaeris*), may weigh up to 60 kg and is found near water across the northeast of South America.

Rodents are an ancient order of mammals, found in the Paleocene of both North America and Asia, indicating an earlier first occurrence. Rodents have always been successful, and myriad fossil forms are found in the Eocene of Asia and North America. Modern rodents are usually divided into three major groups: the cavy-like rodents, or Caviomorpha; the mouse-like rodents, the Myomorpha; and the squirrel-like rodents, the Sciuromoprha. While a convenient division, both paleontological and molecular data suggest that this tri-partite classification may be more useful than it is real.

While supporting a phenomenal diversity in both Recent species and fossil forms, all rodents have adopted a similar jaw structure, diagnostic of the order. The incisors are used for gnawing and clipping and are sharp and ever-growing. Molars are adapted for grinding. The jaw musculature is also modified, with the muscles moving some of their attachments off the zygomatic arch and cranium forward onto the rostrum, with chewing thus providing a forward movement of the jaw.

Rodents have a global distribution and are found in every habitat type from the high Arctic to the driest deserts and the wettest tropical forests. Species like the Norway rat, *Rattus norvegicus*, have hitched rides on ocean-going vessels since people started sailing the seas 10,000 years ago. A crop pest, the Norway rat is responsible for billions of dollars of damage each year. Of greater historical significance, rats have acted as secondary hosts for a number of diseases that plague humans, including the bubonic plague itself, which killed 30 million people in Europe from the fourteenth to seventeenth centuries. Rats have also been one of the main agents of island extinctions.

17. Lagomorpha

Rabbits (Leporidae) and pikas (Ochotonidae), while not diverse when compared to the rodents (80 species in 13 Recent genera), are nearly worldwide in their distribution. Found in both the New World and Old World, rabbits naturally occur on all continents except Australia, where they have been introduced, as they have been to many larger islands around the world. Fossil remains are first found in Paleocene China, and they occur in a wide variety of habitats, from the Arctic snowshoe hare (*Lepus americanus*) to the tropical species of the genus *Nesolagus* (the Sumatran and Annamite rabbits). Rabbits are herbivores, with hypsodont molars, and, like rodents, ever-growing incisors are used to clip vegetation. Body size varies over a relatively narrow range (300 g to 5 kg), and body form is consistent: round heads, large eyes, big ears, and extended, ricochetal hind legs which converge on those of macropod marsupials. Pikas, which are smaller than rabbits (approximately 200 g), closely resemble small caviomorph rodents and are most commonly found on rocky outcrops and talus slopes. Pikas are found in North America along the northwest coast and across the central Asian steppe into Russia.

18. Macroscelidea

The taxonomic position of the elephant shrews is unclear. Similar to kangaroo rats with long noses, with only 15 living species and a poor representation in the fossil record, the order is usually grouped with the lagomorphs and rodents, although earlier taxonomies have suggested an affiliation with insectivores. Found only in Africa, the order has a narrow range of body size, from the tiny 45-g short-eared elephant shrew (*Macroscelides proboscideusi*) to the rather larger 500-g golden-rumped elephant shrew (*Rhynchocyon* *chrysopygus*). Absent in West Africa, elephant shrews occupy a diversity of habitats, from the deserts of Namibia to the lowland forests of central Africa. Despite the diversity of habitats occupied, all species are monogamous, terrestrial, and omnivorous, although insects form a large part of their diets where studied.

IV. EVOLUTIONARY TRENDS

A. Evolution of Brain Size

The defining characters of hominid evolution are an upright gait and an increasing large and complex brain. Humanoids have the largest brain for their body size of any mammal, extant or extinct. Brain size has, not surprisingly, been correlated with intelligence, with other relatively large-brained forms (e.g., dolphins and great apes) being imputed to be more intelligent than the smaller brained mammals. If larger brains do, indeed, confer greater intelligence and survival of an individual is in some way correlated with intelligence, then one would predict strong selection for increasing brain size in mammalian lineages.

At some point in each of their fossil histories, progressive increase in brain size is observed in primates, cetaceans, carnivores, and ungulates. However, increasing brain size has not been a linear effect, but has been punctuated with periods of rapid increase followed by stasis or relatively slow rates of change. Rapid evolution of brain size occurred early in primates, with modern prosimian brain size occurring by the late Eocene. In anthropoids, long held as the best example of progressive evolution of brain size, a rapid increase in brain size in the Oligocene was followed by relative stasis in most lineages, with the exception of the hominid line. Because carnivores, ungulates, primates, and whales have some of the largest brains and are also among the more charismatic species, generalization about the occurrence of progressive brain size evolution in mammals through time may result from a certain large mammal myopathy.

In many lineages of mammal there has been no such progressive increase in brain size—marsupials, edentates, and some lineages of rodents have shown little change in relative brain size since the orders are first seen in the fossil record in the early Tertiary. This stasis is often attributed to differences in predator pressure, with Miocene South American and Australian marsupials not requiring "higher" brain function and hence there having been no evolutionary pressure on brain size. This explanation is unsatisfactory given the high variation in when brain size evolution occurs in different mammalian lineages.

Because brains tend not to shrink in evolutionary time, the variance in brain size will increase with time, and hence any increase in variance, with size bounded at the lower end of a distribution, will result in a larger average brain size. If selection in evolutionary time for increased brain size in different lineages reflects modern patterns of brain size variation, then examining the ecological and social correlates of brain size variation may give us an insight into the selective forces that may have shaped brain size. For instance, growth of the cerebellum has been correlated with locomotion in three dimensions (flight, swimming) as compared to terrestrial motion. Elaboration of the neocortex has been associated with various aspects of learning. Taxa that have prolonged periods of maternal dependence, and presumably long periods of information transfer, have relatively greater neocortex development than taxa with minimal parental association. Within lineages, in monogamous species males and females have similar brain structure, while in those species where males are promiscuous, there is elaboration of the hippocampus, the part of the brain correlated with spatial memory. The hypothesis is that in searching for receptive females, promiscuous males search over large areas, thus requiring greater spatial skills.

B. Cope's Rule

One of the earliest rules applied to mammalogy is Cope's Rule: species within a lineage will show increasingly large body size through evolutionary time. Cope made his findings in the late nineteenth century working on North American fossil assemblages (some of the best preserved of the Tertiary fossil record), yet most studies since Cope have failed to find support for this generalization. Some have argued that because most lineages originate at small body sizes, it is axiomatic that the only place they have to go is up. Others have stated that the pattern seen in the fossil record is a statistical artifact, the result of passive diversification of species within clades rather than the result of any kind of directed evolution. Recently, it was even suggested that while Cope was an advocate of directed evolution, he did not even do any significant analysis of body size trends and that the attribution of the first observation of this phenomenon is misplaced.

As with all paleontological examination, the detection of such trends relies on the quality of the data set used. Most studies of Cope's law have suffered from either a telescoped time frame or data sets that cover a long time run, but focus on a small number of taxa. But a recent study of Cope's law, which used a data set of 1534 species of North American mammals ranging in age from the late Cretaceous to the Pleistocene, suffers from neither of these faults. Species within a genus were followed through their evolutionary histories, and, on average, new species were 9% larger than the older species within the same genus. Diversification in size was not gradual, but changed rapidly at the Cretaceous/Tertiary (K/T) boundary, coincident with the rapid ordinal diversification of mammals. Average body weight of 29 species in the late Cretaceous was 150 g; by the early Tertiary, the average weight of 33 matched species was just over 1000 g, an order of magnitude higher. There is an unambiguous directional trend to larger size in North American mammals, but what could possibly explain this trend?

While data convincingly show that Cope's law is valid, at least for North American taxa, no good explanation has been offered as to why such persistent increases in body size are observed. Not only does body size increase within lineages, but at some point the middle of the size distribution drops out, leaving relatively large and relatively small species. This has led some to suggest that there are, perhaps, optimal body sizes for homeothermic mammals. Scientists have noted in modern assemblages of mammals that there are a disproportionately large number of species that weigh about 100 g (coincident with the average weight of Cretaceous mammals) and that this might be a lower end optimal body size for mammals. This hypothesis is supported by data on the relationship between minimum home range area and body size. While home range size scales roughly with body size in mammals, the relationship is U-shaped, with smaller mammals requiring relatively larger areas than expected. An "optimal" body size is one in which a minimum area is required to support the activities of an individual (the bottom of the U-shaped curve): this minimum occurs between 80 and 200 g.

Upper end optima are not well defined, and gaps in the body size distributions of extant North American mammals are more illusory than real and are easily explained by random statistical variation. If Mesozoic and Cretaceous mammals were competitively excluded by dominant terrestrial vertebrates, the extinction of the dinosaur fauna at the K/T boundary may have opened up the larger end of the body size distribution, allowing mammals to evolve to larger sizes. If larger mammals show higher rates of extinction or lower rates of origination, gaps will develop in the upper end of the mammalian size distribution, perhaps mimicking the random body size gap distribution observed in an ecological time scale and allowing new taxa to continue to evolve to larger sizes.

C. The Island Rule

When mammals colonize islands, a strange thing happens: small mammals, such as rodents, tend to increase in size, while larger mammals, such as carnivores, lagomorphs, and artiodactyls, tend to become smaller. Numerous examples of dwarfism and gigantism have been found in the fossil record and can be documented in extant species, the most notable being the discovery of fossil evidence of pygmy elephants on Mediterranean islands and fossil remains of dwarf mammoth on the Channel Islands of California.

Explanations for this Island Rule have been numerous, and none is completely convincing, but for the most part they focus on a combination of resource limitation, reduced competition as a result of a depauperate fauna, and predator release. For larger species, a limited food supply would favor a smaller body size as individuals of smaller body size require fewer resources, can reproduce more efficiently, and are more likely to leave surviving offspring. Niche partitioning through character displacement will result in selection for the smaller competitor in a feeding guild to become yet smaller. Therefore, it is not surprising that the absence of the larger competitor results in an observed larger body size in the smaller species of a feeding guild. Predation has been hypothesized to have two different effects: in larger mammals, where size is a form of predator defense, release from predators may select for reduced body size. This argument, of course, is confounded with the explanations provided by resource limitation arguments. For smaller mammals, where stealth and crypsis are defense strategies, absence of predators may reduce the adaptive value of small size, thus allowing small mammals to evolve to larger sizes.

A combination of bioenergetics and hypotheses put forth for Cope's Rule may provide a somewhat more synthetic argument for the Island Rule. Studies of mammalian feeding guilds suggest that within a guild of large-bodied mammals, the individuals of smaller species frequently monopolize a greater proportion of the available resources. In guilds dominated by smaller mammals the opposite appears to be true: individuals of larger species within the guild appear to control a greater proportion of the available energy. If there is an optimum size for mammals, arrival on an island in which there was release from predators and guild competitors would allow rapid evolution toward that optimum. Animals that were significantly larger, or smaller, than the optimum would show the greatest divergence between insular and continental body size.

D. These Legs Are for Walking: Predators and Their Prey

The evolution of ungulate locomotion, as typified by the evolution of the equid leg, is a topic covered by every high-school textbook. Cursorial specialization has evolved independently in a number of mammalian lineages, but there is a general pattern of morphological changes that results from a simple calculation: for an animal to move faster, it must increase either the length of its stride or the number of strides it takes. Hence, shifts in morphology must make the leg longer, faster, or both. The lateral reduction and fusing of bones (especially the hand and foot bones, the metacarpals, and metatarsals), the elongation of limbs, the reduction or loss of the clavicle, and the shifting of muscle mass toward the torso with a concomitant increase in the use of tendons to move the limbs are all part of the suite of changes which improve running speed.

Like their herbivore prey, carnivores have also evolved increasingly long-legged, faster forms through time. Carnivores have gone from being short-legged, small creatures to the long-limbed, slender-bodied forms typified by modern pursuit predators such as the cheetah or wolf. These forms appear to have relatively short evolutionary histories, with repeated evolution of similar body types in different carnivore lineages.

The co-evolution of faster herbivores and faster predators has been described as a predator–prey arms race: slight increases in prey speed lead to higher rates of survival of those swift individuals who pass on their genes at disproportionate frequencies. Similarly, faster predators have greater hunting success, produce more offspring, and are better represented in successive generations. There is only one problem with this scenario: the fossil data do not support the hypothesis. Ungulates have tended to evolve ever more cursorial forms throughout the Cenozoic, beginning 55 million years ago in the Eocene. But is was not until the Pliocene, only 5 million years ago, that carnivores "caught up." By this time, most of the "novel" ungulate adaptations for speed were tens of millions of years old.

Long legs confer the ability to move quickly, but what if this ability were a secondary adaptation, the result of selection for another phenomenon? An alternative explanation for cursorial adaptation suggests that the suite of adaptations that has evolved in ungulates has to do with the energetics of movement rather than the energetics of being eaten. All the observed adaptations do one of two things—increase stride length or reduce the costs of moving the leg. Hence, long legs lead to increasing efficiency of movement, whether the animal is walking 10 km to get to a water hole or dashing to get away from a predator. If the opening of the savanna led to increased distance movement in ranging patterns, long, simplified, lightweight legs would confer a great advantage in terms of energy efficiency of movement, whether daily or migratory.

E. Convergent Evolution

Mammalian evolution has been rich with novelty, whether in the evolution of flight in bats, the suite of adaptations evolved by whales that have brought mammals into the sea to live, or more specifically in one lineage the well-developed venom gland found in the hind legs of the platypus. Throughout mammalian evolutionary history, however, some of the more remarkable evolutionary patterns have involved the repeated evolution of derived characters in widely divergent mammal lineages. Convergent evolution spans a wide array of adaptations, from the suite of carnivorous Miocene South American marsupials that resemble extinct and contemporaneous placental carnivores to the striking similarity of burrowing forms in the marsupial mole (family Notoryctidae), the golden moles, (family Chrysochloridae), and true moles (family Talpidae).

Throughout their evolutionary history, mammals have shown trends to increasing specialization in many morphological and correlated ecological functions. For instance, as grasslands became widespread in the Miocene, ungulate tooth morphology shifted from a dominance of low-crowned, or brachyodont, teeth, which are easily ground down through the animal's lifetime, to a dominance of hypsodonty, or high-crown teeth, which last longer when an animal eats siliceous grasses covered in dust and dirt. This evolution has been complemented by an increased complexity of molar teeth to enable herbivores to more thoroughly grind their food. This pattern is observed in both artiodactyls and perrisodactyls. Rodents have solved the problem differently, evolving ever-growing teeth, while proboscideans roll their teeth out of their jaw, each tooth having a limited useful life.

Convergent evolution may address the same problems but find different solutions. The expansion of grasslands and a high-cellulose diet have led to two different solutions to digesting an essentially indigestible substance. Ruminant digestion, widespread in the artiodactyls, uses the fore-gut, or rumen, as a fermentation chamber in which bacteria break down cellulose, making the grass cell contents available for absorption and converting the cellulose into digestible material. Because the passage of materials through the rumen limits intake, artiodactyls tend to be relatively selective, choosing grasses of high quality. In perissodactyls and some kangaroos, a similar solution has evolved—bacterial breakdown of cellulose—but the site of bacterial digestion is in the hindgut and the process is less efficient. While this means that hindgut fermenters cannot draw out as many nutrients from a given pulse of food, by rapid processing of food, hindgut fermenters can gain sustenance from larger volumes of lower quality forage.

Table III lists a few of the more widespread patterns of convergent evolution in mammals.

V. GEOGRAPHY, BIOGEOGRAPHY, AND BIODIVERSITY

A. Gradients of Species Richness: Diversity, Density, Range, and Rapoport's Rule

The tendency for species richness to increase with decreasing latitude was first observed by Alfred Russell Wallace and has been the subject of significant study. Mammals, as well as nearly every taxa studied, appear to show this pattern. While nearly a dozen hypotheses have been suggested to explain the phenomenon, no single explanation appears dominant.

In a study of nearly 200 North and Central American mammals which matched subspecies living at lower and higher latitudes, it was found that those living at lower latitudes had significantly smaller overall latitudinal geographical distributions. The same pattern is observed in a review of 679 North American mammals in which latitude is compared to the range of the species. In this study, in addition to the increase in range size with increasing latitude, a similar pattern is observed in a west-to-east gradient of longitude, with range sizes increasing to the east. Between the tropics and the northern Arctic regions, the average range of a species increases by a factor of 30. Interestingly, this same pattern is not observed in Australian mammals. The continent has a divergent geographic structure, with an arid center fringed by moist mountains, creating habitat diversity (and stability) very different from that observed in northern continents.

The vast majority of North American mammals have a narrow geographical species range that covers but a few habitat types and appear to be habitat specialists. Not surprisingly, patterns of variation are not consistent across orders, with the ranges of carnivores and artiodactyls being larger, on average, and bats and rodents much smaller. Nonetheless, the observed pattern of increasing range with increasing latitude is not an artifact of taxonomy: these patterns are seen both across all mammals and independently within mammalian orders.

Overall, range size shows a log-normal distribution, with very large and very small ranges being exceptional. Species with small ranges would probably result in a greater risk of extinction through severe catastrophic events such as abnormally cold, hot, or dry weather, through disease, or because of predation or competition. Species with unusually large areas may be prone to extinction because of the low densities at which they occur.

It has long been noted that species diversity (or more accurately for mammals, species density, the number of species per unit area) decreases with increasing latitude. But species density also has a strong trend, with species density increasing from east to west. The increase in density is greatest in the more southerly latitudes (30–40°) and correlates with increasingly complex topography, and hence greater habitat diversity.

Is there a connection between these two patterns, with narrowly distributed species packed in more closely at southern latitudes and in more complex habitats? Selection for wide tolerances of temperature and moisture extremes is more likely to occur in the temperate zones, where temperatures can range from freezing or subfreezing in the winter to 35°C in the summer. Patterns of climatic variation are similar over large areas; hence moving from place to place does not usually expose an individual to temperature or moisture extremes that have not previously been experienced. At more northern latitudes, the range of extremes of moisture, temperature, and light become more extreme, and unless a species is migratory, an individual will be exposed to these extremes year after year throughout its lifetime.

While wide climatic tolerance might have few costs, it would confer few benefits in the tropics. In the tropics the range of climatic variation at any given location is low. Microhabitat variation operates over a smaller spatial scale; hence staying in one place does not expose an individual to climatic variation, but moving a few tens of kilometers, or a few hundred meters in elevation, can expose an individual to conditions to which the individual, or its ancestors, have not been exposed.

TABLE III

Examples of Convergent Evolution in Mammals[a]

	Geographic location				
Adaptation	Holarctic	Africa	Madagascar	South America	Australia
Carnivory	Carnivora	Carnivora	Carnivora	Borhyaenidae*	Dasyuroid
	Many families	Many families	Viverrids		Phalangeroid
Pursuit predators	Carnivora	Carnivora		Didelphoid	Dasyuroid
	Dogs	Hyenas		Borhyaenid	*Thylacinus**
	Dog bears*	Cheetah			
	Bear dogs*				
Saber-tooth carnivores	Smilodon*	Felids	Viverrids	Borhyaenid*	Phalangeroid
	Felids		Fossa	*Thylacosmilus**	*Thlacoleo**
	Nimravids*				
Semiaquatic	Insectivora	Insectivora	Insectivora	Didelphoid	Monotreme
	Desmans	Otter shrew	Otter tenrec	Water opossum	Platypus
	Water voles				
	Pantolestids*				
	Rodentia				Rodentia
	Musk rat				Water rat
	Carnivora				
	Otters				
Aquatic	Carnivora				
	Sea lions/seals				
	Cetacea				
	Whales				
Herbivores with cellulose fermentation	Ungulates	Ungulates		Ungulates	Phalangeroids
	Artiodactyla	Proboscidians		Notoungulates*	
	Perissodactyla	Hyracoids		Lipoterns*	
	Some rodents	Arsinotheres*		Pyrotheres*	
		Some rodents			
Hornlike structures	Perissodactyls	Arsinotheres*		Notoungulates*	
	Rhinos			Toxodontids*	
	Brontotheres*				
	Artiodactyls				
	Protoceratids*				
	Oreodonts*				
	Pigs				
	Ruminants				
	Archaic ungulates				
	Uinthatheres*				
	Rodents				
	Mylagaulids*				
Anteaters	Pholidota	Pholidota		Edentates	Dayroids
	Pangolins	Pangolins		Anteaters	Numbat
				Armadillos	Monotremes
					Echidna
Gliding	Rodents	Rodents			Phalangers
	Flying squirrels	Scaly-tailed squirrels			Flying possums
	Dermopterans				
	Flying lemur				
	Paromomyidae*				

[a] From Janis and Damuth (1990). Asterisks indicate extinct taxa.

This has prompted one scientist to note that "mountain passes are higher in the tropics." The physical structure of such passes is, of course, no different in the tropics. But the perception of a dispersing animal trying to cross that pass may vary widely, depending in large part on the way adaptation for extremes in temporal variation of ecological variables has equipped the individual to deal with new microclimates encountered on its journey.

B. Hot Spots of Species Richness and Endemicity

Broader patterns of species diversity and species density may explain large-scale geographic variation in observed diversity of mammals, but within these patterns there are finer grained anomalies in distribution that result in areas with extremely high diversity, so-called hot spots. Across all species of terrestrial vertebrates, 35% of all species are confined to a total of 1.4% of the Earth's land surface. Almost by definition, this includes a large proportion of geographic endemics, and not surprisingly, such hot spots tend to be much more common in the tropics than in temperate areas for reasons obvious from the above discussion of geographical patterns of diversity. This pattern has strong implications for priority setting in conservation of mammals in the tropics, with a greater bang for the buck (as measured by species protected for dollar expended) if conservation investments are focused in these hot spot areas.

The geographical and geological precursors that lead to high levels of species endemism are not well understood. Despite the radical difference in life histories and evolutionary histories, patterns of endemism frequently are highly congruent across taxa, indicating necessary and sufficient conditions are probably common to the evolution of endemicity in mammals and other taxa. Indeed, in Africa and its associated islands, endemism of plants and animals converges, with the hottest of the hot spots for all taxa being found in places like Madagascar, the Karoo, the Cape Province of South Africa, and the Ethiopian Highlands.

The diversification of a small number of families on islands is probably as much due to vicariant or random arrival of animals on the island as it is to any set of geological or climatic conditions. The diverse lemur fauna of Madagascar and the rapid radiation of marsupials on the Australian continent owe as much to isolation as they do to anything else. The rapid, and astonishing, diversification of cichlid fishes in the rift lakes of Africa or the convergent pattern seen in the fish genus *Semionotus* in the Jurassic rift lakes of the East Coast of North America also shows the potential for rapid evolution of endemics in habitat "islands" within a larger continental setting.

However, isolation, per se, does not explain the evolution of endemism on larger continents, where connections to other source populations of mammals have come and gone over the epochs. Africa, unlike other continents, has retained a relatively stable location on the globe through the past 300 million years. Such stability, relative to the drift of other plates, has provided a relatively constant set of ecological conditions as associated with latitudinal changes. Superimposed on this stability, however, have been the making and breaking of connections to Asia, repeated geological upheavals associated with the repeated formation of an African rift, and a cycle of drier periods that led to expansion and contraction of the extent of the central African forests.

Within this geologically unstable environment, there have been areas of remarkable stability, areas that have been consistent in their patterns of temperature and rainfall, whether hot and dry or hot and wet. This climatic stability has led to the evolution of endemic mammals in places as diverse as the Ethiopian Highlands, the forests of the Congo basin (former Zaire), and the Usambara Mountains (although this area is better known for its avian and floristic endemism).

Endemic hot spots appear to be at greater risk than equivalent areas of similar vegetation type. In tropical forests, for instance, in hot spots, only 12% of the original vegetation persists, while 50% of all tropical forests still stand. This suggests that hot spots are an order of magnitude more threatened than the average tropical forest. One explanation for this ties into the evolutionary explanation for hot spots. Many of the same patterns of geography and climatic stability that lead to the evolution of hot spots are also those that lead to the settlement and cultural evolution of humans. Hence, convergent patterns of human cultural diversification and biotic diversification lead to a situation in which people and hot spots are in closer proximity, competing for the same resources and leading to greater conflict than would be expected if species and people were randomly distributed.

VI. NEW DISCOVERIES

A. Patterns of Discovery

The discovery of a new mammal, even a large mammal, is not such an unusual event. In the past 60 years,

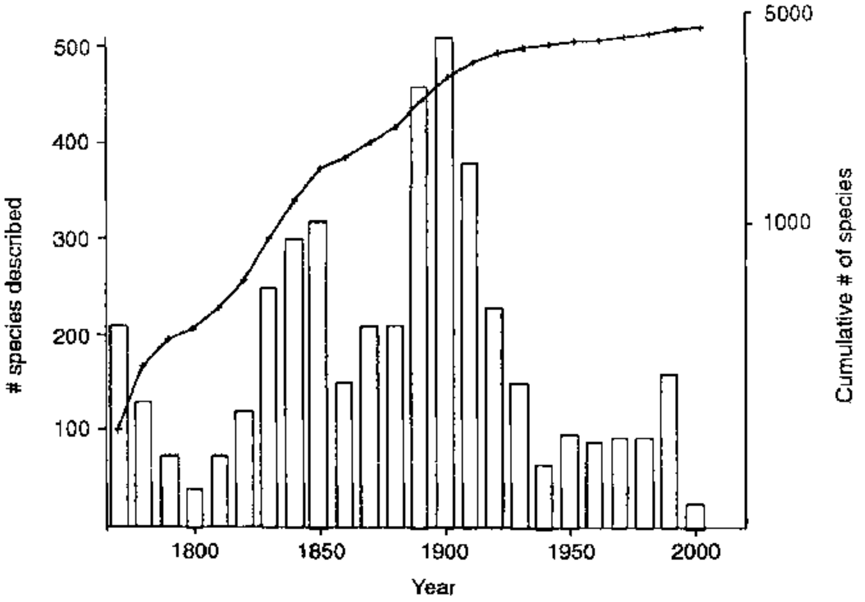


FIGURE 4 The number of new species described (left-hand scale, histogram) and cumulative number of mammal species (right-hand scale, line). After Wilson and Reeder (1999).

17 species of large mammals have been described and several others are known but await formal description. Consideration of the pattern of description of new species of mammals shows that the curve is more or less bell shaped (Fig. 4), with the peak of activity in describing mammals coming early in the twentieth century. While new mammals continue to be discovered each year, the rate at which these discoveries have occurred has not changed significantly in recent decades.

If finding new species is not a particularly rare event, can we predict how many species will eventually be discovered? Recent analysis of the discovery and description of large (>2 m) open-water marine mammals suggests perhaps that 47 species await formal scientific description, that we should continue finding species at a rate of one new species every 5 years, and that we are most likely to find new cetaceans. Given the long and intensive history of commercial whaling, such a finding is surprising.

New species can be named in one of several ways. In some cases, specimens collected many decades earlier were neither properly examined nor adequately described. Hence, when a systematist revises a taxon, "new" species may be found sitting in a museum drawer. Similarly, new forms of evidence from field, laboratory, or genetic studies may force a revision of a species group. Molecular genetics has played a large role in this field, forcing scientists to reassess previously assigned taxonomic names. Finally, previously unknown forms may be found in nature.

B. The Annamites: The Last Frontier for Large Mammals

The Annamite mountain range, which forms the border between the Lao People's Democratic Republic and Vietnam, has yielded a large number of discoveries of new species and rediscoveries of previously described species of large mammals in the past decade. The Vietnamese warty pig, *Sus bucculentus*, long thought to be extinct, was found in a food market in Laos in the early 1990s. The field has been particularly rich for discovery of barking deer, primitive deer found throughout the tropics of southern Asia and Southeast Asia. In 1994, the giant muntjac was discovered in the forests of the Annamites. Large in size and with unusual antlers, the species was thought sufficiently morphologically distinct to deserve its own genus, *Megamuntiacus*. DNA analysis suggested that the species belongs with other muntjacs in the genus *Muntiacus* (*M. vuquangensis*) and not in a genus of its own. Two further Annamite *Muntiacus* species have been found, one a new species from Quang Nam Province, Vietnam (*Muntiacus truongsonensis*), and another from the Lao side of the border, a rediscovery of Roosevelt's muntjac, *M. rooseveltorum*. In neighboring Myanmar (Burma), a 1997 expedition to the far north of the country, an area not visited by scientists since the 1930s, yielded discovery of the world's smallest deer, another species of muntjac, the leaf deer (*M. putaoensis*).

The Annamites have also been the source of two discoveries that have both puzzled and captivated scientists. The first is the saola, or *Pseudoryx nghetinhensis*. The species, which weighs approximately 100 kg, derives both its scientific name and Lao common name from the shape of its horns: long and arced, the species' horns resemble both the Africa oryx and the arc of a lyre-like Lao musical instrument. First found in Vu Quang forest reserve, Vietnam, the species is now known to range widely at higher altitudes in the Annamites. Most closely related to bovids, the saola is clearly unusual and represents a deep branch in the phylogeny of the bovids.

Another, more curious, Annamite discovery was that of a new rabbit species (*Nesolagus timminsii*). The rabbit, which has distinct, dark brown stripes running down both its face and back, a reddish rump, and short ears, was first identified in a food market in Ben Lak, Laos, but since has been photographed in Vietnam. The rabbit's closest relative is a critically endangered species found in Sumatra, about a thousand miles away (*Nesolagus netscheri*). Despite extreme morphological similarity, genetic data suggest that the two species may have diverged about 8 million years ago. The two are so divergent genetically that it was debated as to whether the new (as yet unnamed) species should be placed in its own genus.

C. Cryptic Species, Phylogeography, and Evolutionary Significant Units

New discoveries do not require going to the ends of the earth. The pipistrelle (*Pipistrellus pipistrellus*) is one of the most common bats in Europe, and one of the best studied. In the British Isles it was found that the bat used two frequencies to echolocate, which led scientists to classify these populations as either the 45- or 55-kHz phonic type. Because roosts consisted of bats of a single phonic type, to these scientists it was suggested that the phonic types represented sympatric, but distinct, species. Because of their morphological similarity, they are called cryptic species. Two recent studies have provided further evidence that these two phonic types are, indeed, good species. The two show relatively good separation of diet: while both eat mostly the dipteran suborder Nematocera, the dominant prey groups for the 45-kHz phonic type were in the families Psychodidae, Anisopodidae, and Muscidae, while the families Chironomidae and Ceratopogonidae were the main prey groups of bats of the 55-kHz phonic type. Studies of the cytochrome *b* gene also showed unambiguously that the two phonic types are distinct species. That such cryptic species can be found in such a well-studied group suggests that further genetic studies of mammalian phylogenies will yield some surprises.

Molecular genetics can also be a powerful tool for disentangling the relationship between the geographical and morphological history of subspecies within a species group, the study of phylogeography. Because morphological, geographic, and genetic data frequently tell different stories (as in the striped rabbit example above), an understanding and reconciliation of these data are critical to understanding both the evolution and conservation of mammals. The recent study of Australia's largest extant carnivore, the marsupial tiger quoll (*Dasyurus maculatus*), is instructive. The quoll occurs on Tasmania and on the mainland in two distinct populations, one in the north and one in the south. Previous to genetic analyses, strong body size and morphological convergence led to the southern and Tasmanian populations being grouped as one subspecies (*D. m. maculatus*), while the northern population was classified as a separate subspecies (*D. m. gracilis*).

Genetic evidence suggests, however, that despite their size differences, which may be as much as 50%, the two mainland populations are more closely related to one another, while the Tasmanian population is genetically divergent. Why does this matter? If the two mainland populations constitute one evolutionary significant unit (ESU) while the Tasmanian population constitutes a different ESU, then one management solution would be to manage the mainland and island populations separately. This would ease management concerns as the northern population is classified as endangered, while the southern population is relatively more common. Yet, it is likely that if a female northern quoll were placed in a breeding center with a southern male, because of their great size difference, the male would view the female as dinner, not as a potential mate. Similarly, translocation of southern quolls to the north to augment that population could be disastrous. Hence, while phylogeography suggests one management regime, a more complex solution, with each population managed separately, is recommended.

VII. CURRENT EXTINCTION CRISIS

A. How Many Species Have Gone Extinct?

The current rate of extinction of mammals is, by any measure, frightening. In the past 500 years, approximately 88 species of mammals are thought to have gone extinct. This represents approximately 1.9% of the extant species of mammals. Using the "background" or natural rate of extinction derived from an examination of the fossil record, one would have expected at most one species of mammal to have gone extinct in the same time period. This accelerated rate of extinction can be ascribed, directly or indirectly, to a single cause: humans. While determining what factor is ultimately (as opposed to proximately) involved in a species extinction, habitat loss, introduction of exotic species (including zoonotic diseases), and overharvesting all have contributed to the high historic rates of mammalian extinction.

One would think that assessing the rate of extinction in such a well-studied group as mammals would be relatively easy. Yet, despite relatively good taxonomy and detailed study of the question, there remains considerable debate among mammalogists and conservationists about the number, the precise identity, and the timing of disappearance of those species that may have gone extinct in historical times. Two recent efforts to categorize historical extinctions came up with approximately the same number of species going extinct: 85 by one estimate and 88 by another. What is disconcerting, however, is that the two lists contain only 57 species in common, resulting in, overall, 116 different species being listed by one or another of these studies. This suggests that, even among experts, there is some debate about what actually constitutes an extinction.

How much does this matter when assessing patterns of modern loss of mammalian species? On the one hand, documented extinction rates are so far over background rates that definitive listing will not change the implications for conservation of mammals. On the other hand, such lists are playing an increasingly important role both in setting priorities for conservation (the process of keeping species off the list) and in public debate, and an accurate, defensible list is essential.

A clear ability to agree on what is meant by extinction is critical both to conservation planning and to conservation science. While all differences in lists of extinct species need to be reconciled, much can be learned by analyzing why one set of authors excludes species that other authors list as extinct (Table IV). The reason(s) why authors disagree can tell us much about the source of such errors, or differences of opinion, but also help us focus our efforts more clearly on those species where changes in status affect conservation action.

Most lists examine the extinction of species over a particular time frame. For instance, one might examine mammalian extinctions since the beginning of the age of exploration starting in 1500. The first question one must ask is, Did the species in question really exist at any time during the past 500 years, or was extinction during this time inferred incorrectly either through data in the literature or from misplaced stratigraphy of subfossil material? In Table IV, clearly a plurality of the 76 species excluded were in this category. There is no dispute whether or not these 31 species went extinct, rather just a question of whether their extinction predated the period under study. While inclusion or exclusion of these species will change the calculation of rates of extinction, such changes have little practical applications to conservation.

TABLE IV

A Review of Species Not Accepted As "Extinct"[a]

Species not known to exist in past 500 years	31
Species extant as another taxa	18
Species extinct as another taxa	3
Taxonomy unresolved	3
Species extant	21

[a] Data from McPhee and Flemming (1999).

Similarly, removing or adding a species from a list for reasons of taxonomic uncertainty or taxonomic revision does not manifestly change conservation status of a species. Take for instance the quagga. Distinct from other zebras, with only vestigial stripes, the quagga, *Equus quagga*, was initially thought to be a distinct species of zebra extirpated from the southern tip of Africa at the turn of the twentieth century. Recent molecular analysis of museum specimens showed unambiguously that the quagga was a subspecies of the common zebra, *E. burchelli*—one less extinct species, to be sure—but such revision does not change the way we manage the existing populations of *E. burchelli* Similarly, if two extinct species are found to synonymous, as in the case of Johnson's hutia (*Plagiodontia ipaneum*) and *P. velozi* (no common name), management plans remain unaffected.

When a species thought to be extinct is determined to still be extant, more likely than not the size and status of the extant population will be either totally unknown or known to be critically endangered. Hence, transition from extinct to extant, unlike other categories, has important conservation implications. In Table IV, such cases represent about one-quarter of those species listed. Such a transition should be a red flag indicating that further study, and conservation action, are likely needed.

B. Patterns of Modern Mammalian Extinctions

While the rate of extinction is extremely high in comparison to the expected background rate of extinction, the time frame over which historical extinctions have occurred is so short that it is often difficult to discern patterns in data collected on mammalian extinction. Nonetheless, both taxonomic and geographical patterns do emerge, both of which may be informative if we try to project future patterns of extinction in mammals.

Perhaps the most striking pattern seen when one examines data on historical extinctions of mammals relates to the geographic distribution of extinct species. In the past 500 years, the great majority of extinctions have been on islands, and within island groups, the Caribbean islands have suffered the most extreme loss, accounting for nearly 40% of all recorded extinctions (Table V). Mammalian extinctions are, in this case, no different from those of other well-studied taxa, with over 90% of modern avian extinctions and 89% of modern molluscan extinctions also occurring on islands. In recent times, continental extinctions have, for the most part, been rather rare, and nearly all continental extinctions have been on the continent of Australia.

No sophisticated statistical analysis of the effect of body size and phylogeny has been made of recent extinctions, nor have analyses adjusted extinction rates for extant patterns of diversity. Yet the observed patterns suggest that smaller mammals have been particularly susceptible to extinction. This pattern contrasts sharply with that observed in the fossil record of the Americas during the Pleistocene–Holocene transition, in which the majority of known extinctions were in the megafauna of the region.

Perhaps the most striking set of extinctions is in the Insectivora, where an entire family (Nesophontidae, the Antillean island shrews) has gone extinct. Eleven percent of all recorded mammalian extinctions in the Modern era have occurred in this family. Insectivora, as an order, appears to have a greater propensity to extinction, with 11 species having gone extinct, representing 2.5% of the ±450 described species of Insectivora. This is about 30% above the overall rate of 1.9% for mammals.

Other orders that have overrepresentation in recent extinctions include two eutherian orders, the most diverse being Rodentia (52% of all extinctions, 44% of described mammals). Similarly, among therian mammals, the most diverse order, Diprodontia, with 117 species, which includes possums, cuscuses, wombats, and the koala, shows overrepresentation in the extinction table (6.8% of extinctions, 2.5% of all mammalian species) as does one of the least diverse, the Perameldia, or bandicoots (3.4%/0.45%).

In contrast, while nine species of the order Chiroptera have gone extinct, representing 10% of the known modern extinctions, bats represent nearly 20% of the extant modern mammalian fauna; hence, one could argue bat extinctions have been underrepresented in recent times. Similar arguments could be made for both carnivores (2.2% of extinctions, 5.8% of species) and primates (3.4% of extinctions, 5.1% of species).

TABLE V

Geographic Patterns of Mammalian Extinction[a]

Type	Loss(%)
Islands	
Caribbean	37.5
Pacific	21.6
Indian Ocean	8.0
All other	6.8
Total	73.9
Continental	
Australia	19.3
Africa	4.5
Eurasia	1.1
Americas	1.1
Total	26.0

[a] Data from McPhee and Flemming (1999).

C. Projecting Future Extinctions

Compiling lists of recently extinct species and studying current extinctions tells us something of the recent history of changes in mammalian biodiversity. The strength of such lists is that they are an assessment of the global patterns of extinction and tell us something about the persistence of a species, family, order, or size class. To better understand the immediate future of mammalian extinctions (perhaps the next 200 years) rather than reflect on what has gone extinct, there is greater value in examining which species, families, and orders are under threat and discussing how threats will lead to decline, and ultimately to extinction. These processes are more fully discussed elsewhere in this encyclopedia, as are patterns for particular orders of mammals.

For mammals overall, however, patterns of threat do not correlate well with patterns of recent extinction. A recent assessment of threatened species by the World Conservation Union (IUCN) suggests that 25% of extant mammal species are threatened with extinction. Some of the groups that are overrepresented in analyses of recent extinctions are prominent in the list of threatened species: rodents are somewhat underrepresented, with "only" 17% of the species in the order threatened with extinction; insectivores, however, look likely to dominate the lists of extinct species in centuries to come, with 36% of the species in this taxon threatened. What is more interesting is that some orders that have shown relatively low rates of extinction such as carnivores and primates have high proportions of their taxa threatened. In carnivores nearly 26% of all species are under threat, while nearly half the primates, 46%, may go extinct in the next century.

Analysis of the extinction of subspecies or populations of animals gives some insight into what is generating this pattern. Among both carnivores and primates, rare species have been subject to increasing threats from habitat loss and fragmentation, isolating endemics in smaller and smaller patches. But wide-ranging species are also showing promise for future extinction through similar processes acting at larger scales. The spotted hyena, while still abundant and occurring at relatively high densities for a large carnivore, requires a matrix of resources that are widely dispersed across a landscape. Combined with a relatively low level of dispersal, habitat fragmentation and isolation suggest the species will come under increasing threat. The African wild dog still has a nearly continental distribution, but most populations are small (fewer than 100 individuals) and increasingly isolated. Threats to the species amplify one another: habitat loss and fragmentation, loss of prey, disease introduced by domestic companion animals, active hunting, and road kills all contribute to mortality. The decline of the species in West Africa, where populations became isolated and then went extinct one by one, like candles being snuffed, is probably indicative of patterns that are evolving in central and East Africa. Increasingly, while the species is still extant, and will remain so for decades in southern Africa, it is being lost as a component of the great majority of ecosystems in which it once lived.

While extremely valuable both for planning and for developing an understanding of which species are most likely to go extinct, lists of extinct or threatened mammals only address the concern of evolutionary extinction, the ultimate loss of a species. Increasingly, scientists who study mammals, and lobby for the preservation of mammalian diversity in particular and biodiversity more widely, will have to address issues related to the local, regional, and global ecological consequences of a species march toward extinction.

See Also the Following Articles

BIRDS, BIODIVERSITY OF • ENDEMISM • EXTINCTIONS, MODERN EXAMPLES OF • HOTSPOTS • LATITUDE, COMMON TRENDS WITHIN • MAMMALS, CONSERVATION EFFORTS FOR • MAMMALS, LATE QUATENARY, EXTINCTIONS OF • MAMMALS, PRE-QUATENARY, EXTINCTIONS OF • VERTEBRATES, OVERVIEW

Bibliography

Alroy, J. (1998). Cope's rule and the dynamics of body mass evolution in North American fossil mammals. *Science* 280, 731–734.

Alroy, J. (1999). The fossil record of North American mammals: Evidence for a Paleocene evolutionary radiation. *System. Biol.* 48, 107–118.

Baillie, J., and Groombridge, B. (1996). *1996 IUCN Red List of Threatened Animals.* IUCN, The World Conservation Union, Gland, Switzerland.

Barlow, K. E. (1997). The diets of two phonic types of the bat *Pipistrellus pipistrellus* in Britain. *J. Zool.* 243, 597–609.

Carroll, R. L. (1988). *Vertebrate Paleontology and Evolution.* W. H. Freeman, New York.

Dayan, T., and Simberloff, D. (1998). Size patterns among competitors: Ecological character displacement and character release in mammals, with special reference to island populations. *Mamm. Rev.* 28, 99–124.

Firestone, K. B., Elphinstone, M. S., Sherwin, W. B. V., and Houlden, B. A. (1999). Phylogeographical population structure of tiger quolls *Dasyurus maculatus* (Dasyuridae: Marsupialia), an endangered carnivorous mammal. *Mol. Ecol.* 8, 1613–1625.

Janis, C. (1994). Do legs support the arms race in mammalian predatory/prey relationships? In *Vertebrate Behavior As Derived from the Fossil Record* (J. Horner and L. Ellis, Eds.). Columbia Univ. Press, New York.

Janis, C., and Wilhelm, P. B. (1993). Were there mammalian pursuit predators in the Tertiary? Dances with wolf avatars. *J. Mamm. Evol.* 1, 103–125.

Janis, C. M., and Damuth, J. (1990). Mammals. In *Evolutionary Trends* (K. J. McNamara, Ed.), pp. 301–346. Univ. of Arizona Press, Tucson.

Jones, G., and Vanparijs, S. M. (1993). Bimodal echolocation in pipistrelle bats: Are cryptic species present? *Proc. R. Soc. London, Ser. B: Biol. Sci.* 251, 119–125.

Kingdon, J. (1989). *Island Africa.* Princeton Univ. Press, Princeton, NJ.

Macdonald, D. W. (Ed.) (2001) *Encyclopedia of Mammals*, 2nd ed. Oxford Univ. Press, Oxford.

McKenna, M. C., and Bell, S. K. (1997). *Classification of Mammals above the Species Level.* Columbia Univ. Press, New York.

McPhee, R. D. E., and Flemming, C. (1999). *Requiem: Aeternam:* The last five hundred years of mammalian extinctions. In *Extinctions in Near Time* (R. D. E. MacPhee, Ed.), pp. 333–371. Kluwer Academic/Plenum, New York.

Morell, V. (1996). New mammals discovered by new explorers. *Science* 273, 1491.

Novacek, M. J. (1992). Mammalian phylogeny: Shaking the tree. *Nature* 356, 121–125.

Novacek, M. J., Rougier, G. W., Wible, J. R., McKenna, M. C., Dashzeveg, D., and Horovitz, I. (1997). Epipubic bones in eutherian mammals from the late Cretaceous of Mongolia. *Nature* 389(6650), 483–486.

Pagel, M. D., May, R. M., and Collie, A. R. (1991). Ecological aspects of the geographical distribution of mammalian species. *Am. Nat.* 137, 791–815.

Raven, P. H., and Wilson, E. O. (1992). A fifty-year plan for biodiversity surveys. *Science* 258, 1099–1100.

Sharman, G. B. (1970). Reproductive physiology of marsupials. *Science* 167, 1221–1228.

Surridge, A. K., Timmins, R. J., Hewitt, G. M., and Bell, D. J. (1999). Striped rabbits in Southeast Asia. *Nature* 400, 726–726.

Van Valkenburgh, B. (1999). Major patterns in the history of carnivorous mammals. *Ann. Rev. Earth Planetary Sci.* 27, 463–493.

Vaughan, T. A., Ryan, J. M., and Czaplewski, N. J. (2000). *Mammalogy*, 4th ed. Saunders College Publishing, Philadelphia.

Wilson, D. E., and Reeder, D. M. (1999). *Mammal Species of the World*, 2nd ed. Smithsonian Institution Press, Washington, D.C.

MAMMALS, CONSERVATION EFFORTS FOR

E. J. Milner-Gulland* **and R. Woodroffe**†
**Imperial College, London and †University of Warwick, United Kingdom*

GLOSSARY

conservation efforts Any action that aims to reduce the probability of extinction of a taxon over a specified time period.
***ex situ* conservation** Conservation activities that involve individuals held outside their native habitats (e.g., in zoos or seed banks).
***in situ* conservation** Activities which aim to conserve wild populations in their native habitats.
mammal Member of the order Mammalia. A species that provides milk for its young and has fur. They typically (but not exclusively) bear live young.
persecution Deliberate killing of animals perceived to be a nuisance.
sustainable use Exploitation of wildlife in a manner that avoids depletion of the resource (e.g., limited hunting, ecotourism).
terrestrial A species that lives on land (as opposed to marine or fresh water) for the majority of its life cycle.

CONSERVATION EFFORTS for terrestrial mammals must start with a consideration of the particular threats that face these species. Terrestrial mammals are a disparate group, but similarities in their conservation needs can be discerned, particularly when considering large mammals. Large-bodied species tend to be vulnerable to extinction both because of their biology and their interactions with humans. They tend to need large areas for survival and to be relatively slow growing, making their populations less resilient. They also tend to be more likely to be killed by humans for their meat, trophies, or because they are a danger or a nuisance. Because of these varied threats, conservation efforts for terrestrial mammals are broader than for some other taxa; they encompass both habitat protection and direct protection. Conservation efforts for terrestrial mammals are also noteworthy because they are often treated as flagship species for the rest of biodiversity conservation, sometimes at the expense of other species. This makes them important more broadly for conservation policy.

I. INTRODUCTION

In discussing conservation efforts for terrestrial mammals, we first need to identify what is particular about their conservation needs. It is hard to generalize about such a broad range of species, from bats and mice through antelopes and lynx to elephants, but there are some common characteristics that predispose terrestrial mammals to threat, at least among larger-bodied species. These characteristics relate both to natural history and to human attitudes.

Large mammals have high energetic demands. They grow slowly, range widely, and occur at low population densities. Together these linked characteristics mean that large mammals are vulnerable. Their population dynamics are less resilient than those of other species, since they cannot bounce back quickly from population declines. Furthermore, they are more vulnerable to habitat loss than other species, because they rely on larger areas of habitat.

The natural history of mammals is not the only feature that predisposes them to vulnerability—large mammals are often under threat because they are worth killing. Being large bodied, most are valuable for meat; those that have trophies such as horns, tusks, or furs can be extremely valuable. Large mammals can also be prone to human persecution because they represent a threat, both to humans themselves and to their livelihoods, as predators on livestock or competitors for land and resources.

Large mammals can have a significant impact on their environments. This creates an intimate association between the conservation of terrestrial mammals and the conservation of other species in their ecosystems; conservation efforts targeted toward mammals might, therefore, have added conservation value over and above the individual species being conserved. Some such species have a profound impact on the ecosystems they inhabit. For example, the destructive effect of elephants (*Loxodonta africana*) on vegetation means that a trend in plant species richness with elephant density is visible; in the 1980s elephant densities inside Amboseli National Park, Kenya, were very high as they sheltered from poachers in the relative safety of the park. Numbers were very low outside the park, producing a gradient of elephant density that could be measured from damage to Acacia trees, a favorite food plant. Species richness of all plants (whether or not eaten by elephants) was highest at the boundary of the Park, at intermediate elephant densities (Fig. 1). Likewise, predators may have a marked effect on community structure. For example, local extinction of coyotes (*Canis latrans*), in parts of Southern California was associated with disappearance of a variety of smaller species (Crooks and Soulé, 1999), yet the distribution and densities of coyotes themselves have been shaped by the recent decline of a still-larger predator, the wolf (*Canis lupus*).

In this chapter, we shall first discuss in general terms some of the threats facing terrestrial mammals and

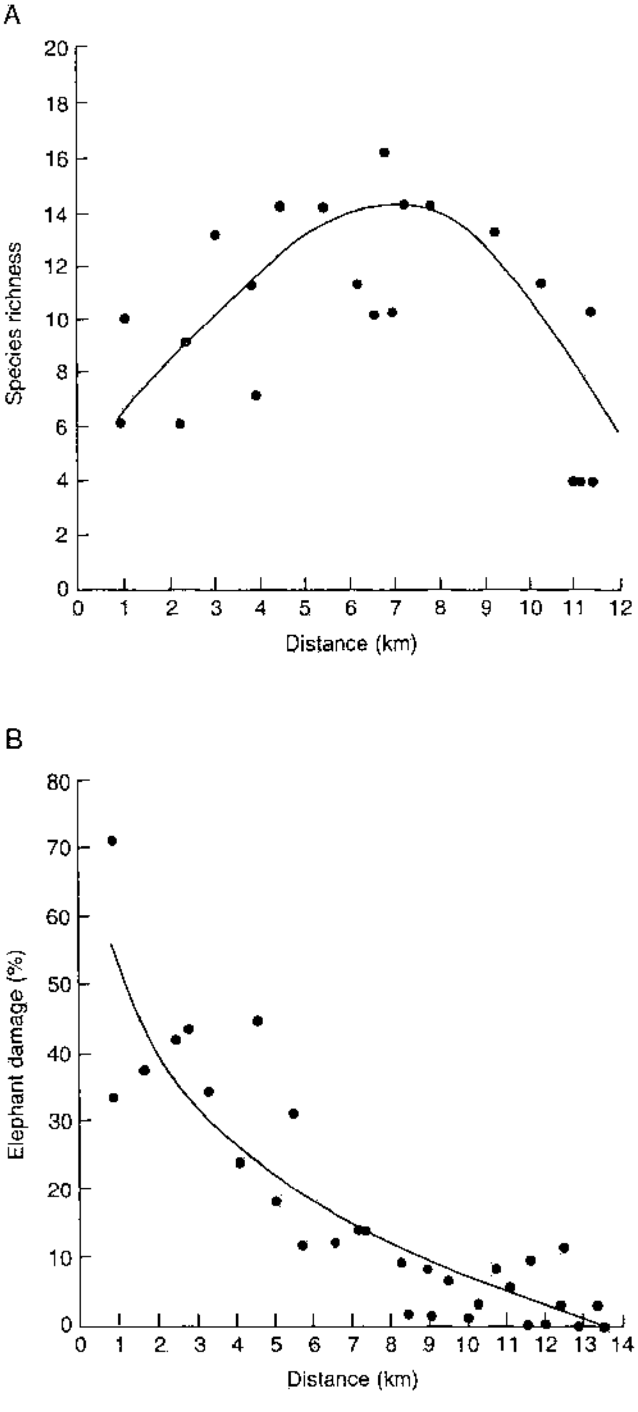


FIGURE 1 (a) Plant species richness along a 12 km transect from the center of Amboseli National Park, Kenya. The Park boundary is at 6 km. (b) Changes in elephant density along the transect from the center of the park, measured by elephant damage to acacia trees. From Western, D. (1989), The ecological value of elephants: A keystone role in Africa's ecosystems. In *The Ivory trade and the future of the African Elephant*, Ivory Trade Review Group, Queen Elizabeth House, Oxford, UK.

which species are particularly vulnerable to these threats. We shall then discuss conservation measures that have been used to address some of the threats. Finally, we shall use in-depth case studies to examine how conservation efforts have been targeted at particular species, concluding by placing terrestrial mammal conservation into the broader context of international conservation efforts.

II. THREATS TO TERRESTRIAL MAMMALS

Like other taxa, many terrestrial mammals are at risk from habitat loss. As people become more numerous, they modify their surroundings in a variety of ways; this process creates new environments hostile to many wild species. Direct conversion of natural habitat threatens the high proportion of mammal species that cannot survive in farmed or urban environments. Many terrestrial mammals' large body size and high energy requirements make them particularly sensitive to human activities.

Direct killing by people is a serious threat, which perhaps affects terrestrial mammals more than other groups (with the possible exception of marine mammals and commercial fish species). Some forms of killing involve exploitation. Subsistence hunting of deer, antelopes, and many primate species can cause population declines, as can commercial hunting to fuel local and international trade. The bushmeat trade (trade in meat from wildlife for local consumption) is predominately based on mammal species; recent estimates have put the annual offtake for this trade at 1000 tonnes of meat a year in Central Africa alone (Robinson *et al.*, 1999). There is a strong interaction between hunting and habitat loss; as an area becomes more accessible through the creation of logging roads, for example, hunting rates increase. Hunting for the international trade has caused some of the most serious and well-publicized species declines and includes hunting of elephants to make ivory carvings and spotted cats to make fur coats. Particularly insidious is the demand for animal body parts used in traditional Asian medicines; this has caused declines of tigers (*Panthera tigris*, in demand for their bones) and bears (Ursidae, in demand for their gallbladders).

Not all killing of mammals by people constitutes exploitation; several species, especially the large carnivores, have suffered serious declines through persecution. Large-bodied species such as elephants and bears are a risk to human life and are rarely tolerated in areas with high human densities. Much of the problem with human-elephant relationships in recent years is concerned less with ivory than with the fact that elephants are dangerous neighbors. Elephants can flatten a crop in minutes and are responsible for large numbers of human injuries and fatalities each year. Likewise, large carnivores are feared and perceived as a threat to livestock and have been the victims of organized eradication campaigns that saw, for example, the extirpation of wolves from most of the coterminous United States, and African wild dogs (*Lycaon pictus*) across much of Africa, in the first half of the 20th century.

Humans may also cause inadvertent declines through exploitation or control of relatively common species. Traps set to catch rabbits (*Oryctolagus cuniculus*) are a serious threat to endangered Iberian lynx (*Lynx pardina*), which are accidentally captured and killed. Likewise giant panda (*Ailuropoda melanoleuca*) populations are seriously affected by accidental capture in snares set for other species. More seriously still, extermination campaigns for prairie dogs (*Cynomys* spp.) inadvertently caused the collapse of black-footed ferret (*Mustela nigripes*) populations.

Human activities also influence wild mammals' exposure to "natural" threats such as disease and predation. Frequent local extinctions of bighorn sheep (*Ovis canadensis*) in the Western United States, leading to serious species decline, have been largely attributable to infectious diseases contracted through contact with domestic sheep. Small, isolated populations are especially vulnerable to the threat of disease; thus disease may deliver the final *coup de grace* to species already crippled by other factors. Likewise, these populations may be unable to cope with predation pressures readily sustained by larger populations. When exotic predators or pathogens are introduced to naïve species, the impact can be more severe still. The introduction of red foxes (*Vulpes vulpes*) to Australia decimated populations of the smaller marsupial species, and the introduction of rinderpest—a viral disease of Asian cattle—to Africa caused widespread and often irreversible declines of many African ungulates.

The recent updating of the World Conservation Union's (IUCN) Red Lists of threatened species provides an opportunity to analyze the relative importance of different threatening processes. The IUCN categories of threat have limitations, reflecting experts' subjective assessments of which threatening processes are the most important for a species. They are also biased in that poorly known or unknown taxa are not well represented. However, mammals are relatively well studied in comparison with other taxa, so that for them, the IUCN categories of threat can be used as a basis for assessing the relative importance of various threatening processes. This has been done by Mace and Balmford (in press), who show that about 25% of mammal species are threatened. This is higher than the threat levels others have estimated for birds, but not as high as for some other taxa. Mace and Balmford identify five major threat types: habitat loss makes up 47% of the threats to mammals, overexploitation makes up 34% of threats (66% of which is local exploitation, 21% commercial exploitation, 5% by catch and incidental catch, and 8% persecution and hunting as a pest). Other threats, such as introductions (13%) and rarity (6%), are less common among mammals. Thus any analysis of the threats to mammals and the conservation efforts needed to counteract them is likely to emphasize habitat loss and overexploitation.

Another useful technique for looking at the effect of threatening factors on extinction risk compares the characteristics of populations that have become extinct with those that survive. These have shown the following links:

Vulnerability correlates with species' natural history. Among carnivores, vulnerability is clearly associated with the extent of a population's contact with people. For example, carnivores with large home ranges frequently travel beyond the borders of protected area, where they come into contact with people. Such contact is often fatal, with the result that wide-ranging species require larger parks and reserves for effective protection (Woodroffe and Ginsberg, 1998).

Vulnerability also correlates with degree of human use of the mammal's habitat. The direct and indirect effects of human activity on mammals lead to general associations between high human population density and local extinction of mammal populations. Elephants, primates, large and small carnivores all demonstrate such associations. It is not clear to what extent these associations reflect direct killing of mammals by people, destruction of mammals' habitat by people, or, on the small scale, simple avoidance of people by mammals.

Individual species may be threatened by a number of different factors, so that the type of conservation effort needed can vary widely between species. However, it is possible to discern some broad patterns that can help to identify species that may be at risk of extinction. These patterns relate both to the biology of the species and to the level and type of human pressure on the species and its habitat.

III. TYPES OF CONSERVATION EFFORT

The threats described here demand a variety of conservation measures, many of which have been applied to terrestrial mammals. We discuss these measures briefly, giving more detailed evaluations of their costs and benefits in the case studies.

Legal protection of areas is the most traditional form of conservation; this approach has the advantage of protecting habitats, as well as particular species, against a multitude of threats associated with human activities. While reserves are often targeted at a particular species, and even named as such (e.g., Addo Elephant Park, Gemsbok National Park), well-planned reserves can protect multiple species, habitats, and landscapes. Large mammals are often used as "flagships" to promote such reserves (e.g., India's Project Tiger). Despite their effectiveness at protecting some species, limitations on local people's rights to inhabit or exploit such areas have attracted controversy and may undermine the longer-term sustainability of reserves.

Legal protection of species or populations may also be used as a conservation measure. For example, stringent legislation in the United Kingdom protects the European badger (*Meles meles*) from killing and disturbance. Difficulties of enforcement mean that such legislation may have limited value as a conservation tool, however; for example, babirusa wild pigs (*Babirusa babyroussa*) are still openly traded for meat in Indonesia despite having full legal protection. Species threatened by exploitation may be protected by *legislation governing trade* rather than preventing killing. For example, in the Gambia it is not illegal to kill leopards, but international trade in their skins would be prohibited through the Convention on International Trade in Endangered Species (CITES).

Problems surrounding the enforcement—and natural justice—of legislation prohibiting the exploitation of wild populations or use of protected areas have led to initiatives aimed at conservation through *sustainable use*. The rationale behind such initiatives is that sustainable exploitation, with the benefits accruing to local communities, should ensure that local people value and protect natural resources. Like the establishment of reserves, this approach has attracted controversy; the array of political, legislative, and logistical concerns that it entails is different from, but no less complex than, that surrounding more traditional legal protection.

Threatened populations may also be protected through direct management. *Habitat management* involves measures such as provision of waterholes or den sites, or control of predators or prey. *Population management* describes a variety of interventions on the population being conserved, such as vaccination against infectious diseases and supplementation from other populations, wild or captive. Species conservation may also involve reestablishment of extinct populations through reintroduction of captive-bred or wild-caught mammals. Such measures have been important in the recovery of some endangered species (e.g., golden lion tamarins, *Leontopithecus rosalia*, red wolves, *Canis rufus*), as well as the reestablishment of populations of species that remain widespread but have suffered local extinctions (e.g., gray wolves).

Finally, the conservation of some endangered species may benefit from *captive breeding*, maintaining a population in captivity that may or may not be intended for release into the wild. Such efforts have proven valuable for critically endangered species reduced to their last few individuals (e.g., Arabian oryx, *Oryx leucopus*, black-footed ferret), in which managed breeding can minimize the loss of genetic variation.

IV. CASE STUDIES

A. African Wild Dogs

African wild dogs provide a case study of the interactions between persecution and habitat fragmentation in leading to a threat of extinction, and this example illustrates how, once depleted, a species can become more susceptible to other factors such as disease and competition with other species. Current conservation efforts for this species include protection in large reserves and reintroductions. We critically examine these efforts and suggest others that may also be useful.

1. Background

The African wild dog, *Lycaon pictus* (Fig. 2), is one of the world's most endangered carnivores. Formerly widespread throughout sub-Saharan Africa, wild dogs have disappeared from 25 of the 39 countries in which they were formerly found, leaving a wild population of around 5000 animals (Fig. 3; Woodroffe *et al.*, 1997). The species' extreme rarity is illustrated by the fact that Africa's elephants currently outnumber its wild dogs by about 100 to 1.

FIGURE 2 Picture of a wild dog. African wild dogs live in close-knit social groups but their overall population density is low and their home ranges are enormous.

2. Threats

Why have wild dogs populations collapsed so dramatically, when other large carnivores have experienced less serious declines? The answer probably reflects both the

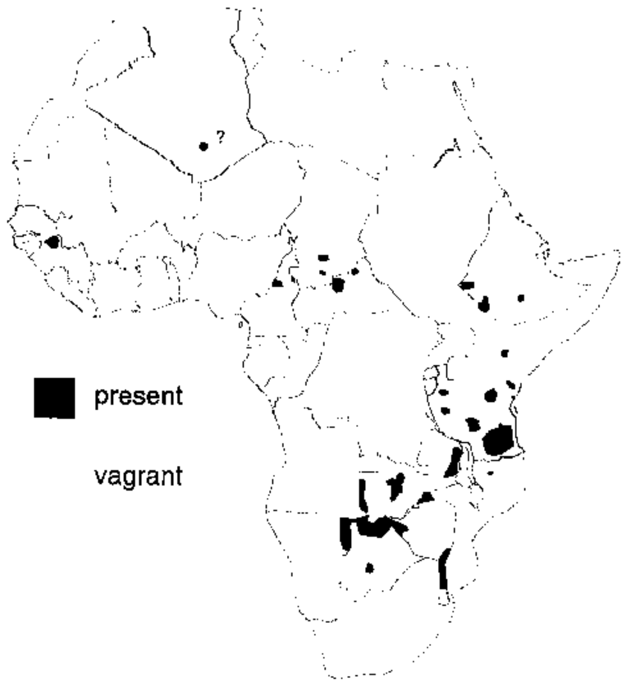


FIGURE 3 Distribution of wild dogs *circa* 1995. Wild dogs' historical distribution covered the majority of sub-Saharan Africa. The population marked with a question mark (?) is known only from unconfirmed reports. Data are from Woodroffe *et al.* (1997).

particularly intense persecution that wild dogs have suffered from people and wild dogs' unusual natural history, which seems to have made them especially sensitive to human activities.

The decline of wild dogs can be linked to the growth and expansion of human populations—they persist only in countries and areas with low human densities (Fig. 2). Because they occasionally kill livestock, wild dogs have long been shot and poisoned by farmers anxious to protect their stock. In the earlier parts of the 20th century, however, wild dogs were the victims of organized eradication campaigns in many parts of Africa. In Zimbabwe, for example, vermin control units shot 2430 wild dogs between 1956 and 1960, and the species was classified as vermin until 1975 (Childes, 1988). Control programs were not intended only to protect livestock. Wild dogs were believed to "[kill] wantonly far more than they need for food, and by methods of the utmost cruelty," so that "When the Uganda national parks were established it was considered necessary, as it had often been elsewhere, to shoot wild dogs in order to give the antelope opportunity to develop their optimum numbers" (Bere, 1955). Wild dogs were shot in Serengeti National Park, Tanzania, as recently as 1973, and in Niger until 1979 (Woodroffe *et al.*, 1997). Thus wild dogs experienced human-caused declines in areas where other species were protected.

Two lines of evidence suggest that such organized persecution is not the sole explanation for wild dogs' ongoing decline. First, wild dogs continue to disappear from protected areas, even though persecution inside reserves halted more than 20 years ago. Second, wild dogs were not the only large carnivore subject to organized control. Other species (e.g., spotted hyenas, *Crocuta crocuta*) were equally reviled and persecuted with similar enthusiasm, but persisted relatively well both inside and outside protected areas (Woodroffe, in press).

Wild dog ecology and behavior appears to have made them particularly sensitive to the persecution and habitat fragmentation associated with the growth of human populations. Even where they are protected, wild dogs live at extremely low population densities (average circa 2 adults per 100 km^2) and range very widely (average circa 600 km^2 per pack). This appears to be a behavioral response to predation by lions (*Panthera leo*), which are a major cause of mortality for both adults and pups. Wild dogs avoid their larger competitor, favoring areas of low prey density where lions are scarce; thus wild dogs effectively experience low prey density even where prey are abundant, and wild dogs are commonest where lions are scarce. Wild dogs' low density means that very large areas ($\geq$10,000 km^2) are needed to support viable populations (Woodroffe *et al.*, 1997). Worse still, their wide-ranging behavior means that few packs remain entirely within the borders of protected areas. Shooting, poisoning, accidental snaring, and road accidents—most of which occur on or outside reserve borders—account for more than half of the adult mortality recorded by studies of protected populations (Woodroffe *et al.*, 1997). Ranging beyond reserve borders also increases contact with domestic dogs, a reservoir of infectious diseases threatening wild dogs (Woodroffe *et al.*, 1997). Such edge-related mortality is most severe where substantial human populations abut reserve borders, and it is powerful enough to cause extinction in small reserves with high perimeter-to-area ratios (Woodroffe and Ginsberg, 1998). Thus, wild dog populations inhabiting small reserves ($\leq$5,000 km^2) face a double jeopardy: small reserve size and high mortality due to strong edge effects. Wild dogs' behavioral response to competition with lions seems to predispose them to local extinction in fragmented habitats.

3. Conservation Efforts

Thus far, few *in situ* conservation measures have been implemented specifically to protect wild dogs. The species has persisted primarily in very large protected areas (Woodroffe and Ginsberg, 1998), and the maintenance of such areas has the highest priority for continent-wide wild dog conservation (Woodroffe *et al.*, 1997). As Africa's growing human population makes increasing demands on the landscape, and reserves are de-gazetted, this is certainly not a trivial recommendation. Any measures that expand the area of contiguous land available to wild dogs will be beneficial; establishing cross-border parks and buffer zones and encouraging land use such as game ranching on lands bordering reserves will all contribute to wild dog conservation. Such landscape-level management would conserve wild dogs by

1. maximizing the size of wild dog populations, making them more resilient to perturbations such as disease outbreaks,
2. ensuring that wild dogs' prey base is maintained, and
3. minimizing the proportion of packs exposed to human activity by lowering perimeter-to-area ratios of wildlife areas.

Few of Africa's reserves exceed the 10,000 km^2 minimum needed to provide adequate protection for wild dog populations large enough to remain viable. Populations remaining in smaller reserves, or outside protected areas, are likely to require active population management to minimize mortality due to persecution, accidental snaring, road accidents and disease (Woodroffe *et al.*, 1997). Such measures will demand

1. working with local farmers to devise husbandry methods that effectively protect livestock from wild dogs, using the findings as the basis for local education campaigns,
2. reducing snaring in protected areas through intensified antipoaching activities,
3. strict limitation on the construction or improvement of roads in and around protected areas, and
4. controlling the growth of domestic dog populations around wildlife areas to limit the spread of infectious disease.

Reintroduction has been advocated as a means of conserving wild dogs. However, although wild dogs have been extirpated across large tracts of Africa, especially in West and Central Africa, it is unlikely that reintroduction has any role to play in restocking these areas. Trials in southern Africa have shown that reintroduction is technically possible, but there are no suitable reintroduction sites in west or central Africa, and few in eastern and southern Africa; candidate sites are too small, or too poorly protected, or entail uncontrolled disease problems (Woodroffe *et al.*, 1997). Furthermore, no wild dogs with the appropriate west, central, or east African genotypes are available for release in the areas where reintroduction is most needed (Woodroffe *et al.*, 1997). Plans are underway to establish a "managed metapopulation" of wild dogs in small, fenced reserves scattered across South Africa's highly fragmented landscape, to be maintained by frequent translocation of animals between sites. This approach is locally valuable but will not establish a population likely to remain viable without intensive management in perpetuity. For these reasons, protection of remaining wild dog populations—and the vast landscapes required to support them—currently represents a better investment than any attempt at reintroduction.

B. Rhinoceroses

Rhinos are threatened predominately by the lucrative international trade in their horns. Conservation efforts have focused on strong protection in small areas, leading to the species existing in increasingly unnatural conditions. Strong laws have also been implemented, banning the commercial international trade in rhino products, although there is now a move toward conservation through trade, facilitated by dehorning. Captive breeding is also an important component of conservation efforts for some species.

1. Background

There are five species of rhino, spread across Asia and Africa. None of the species is entirely secure. The Sumatran (*Dicerorhinus sumatrensis*) and Javan (*Rhinoceros sondaicus*) rhinos are on the brink of extinction in the wild, with population estimates of 300 and <70, respectively (Foose and van Strien, 1998). The Indian rhino (*Rhinoceros unicornis*) population is around 1600 (Martin, 1996), with strongholds in the Royal Chitwan National Park, Nepal, and Kaziranga National Park, India. In Africa, there was a reversal of fortunes in the 1980s and 1990s; previously the black rhino (*Diceros bicornis*) was relatively widespread, with a 1970 population estimated in the high tens of thousands. By 1987 the population estimate had dropped to 3800, and the 1995 estimate was 2400 (Brooks, 1996). These estimates are disputed, but the rapid decline is clear. White rhinos (*Ceratotherium simum*), however, were already at very low numbers by 1900, and have only recently stabilized and started to increase under strong conservation protection. A recent population estimate was around 7500 individuals (Brooks, 1996). There are two subspecies of white rhino; the southern white rhino is currently secure, but the northern white rhino is confined to a population of a few tens of individuals in the Garamba National Park, Democratic Republic of Congo; this population is entirely reliant on continued protection.

2. Threats

Rhino horn has a long history of use and value. Ctesias of Knidos, the Persian court physician from 416 to 398 B.C., wrote of the value of Indian rhino horn drinking cups as poison detectors. Asian rhinos had already been heavily hunted by the beginning of the 19th century, to satisfy demands for horn in the Asian traditional medicine markets (Leader-Williams, 1992). In the 19th century, European game hunters targeted African rhinos along with other big game animals. In the 1830s, Cornwallis Harris saw white rhinos "in almost incredible numbers" in the Cape of South Africa (Martin and Martin, 1982), while Selous (1908) commented on the thousands of white rhinos being killed by hunters in the 1870s, and by 1899 was writing that "the two white rhinos which I shot in 1882 are the last of their species that I have ever seen alive . . . , and when I left Africa towards the end of 1892, I fully expected that these animals would become extinct within a short time."

Wars and the European withdrawal from Africa reduced hunting pressure in the first half of the 20th century, but the rhino horn trade increased to high levels in the 1970s. Oil revenues in Yemen led to large-scale imports of horn in the mid-1970s for dagger handles, while increasing prosperity in the Far East led to an increase in demand for rhino horn for medicinal use (Leader-Williams, 1992). Contrary to popular myth, rhino horn is not primarily used as an aphrodisiac, but as an anti-fever drug. CITES came into force in 1976, prohibiting the international commercial trade in rhino horn; however, key horn consumers such as Japan and Yemen did not ban horn imports until the 1980s.

Although total volumes of rhino horn traded might be significantly lower since the CITES ban, the populations from which it originates are now so small that any level of poaching is a threat to their survival. Thus the 1980s and early 1990s saw a battle between conservationists attempting to protect rhino populations through aggressive law enforcement and poachers killing rhinos for the lucrative illegal horn trade. Several authors have demonstrated a link between the amount of investment in law enforcement and a country's success in protecting their rhinos (e.g., Martin, 1996). Hwange National Park, Zimbabwe, was extremely successful at protecting its rhino population until a 4-month hiatus in funding led to the entire rhino population being wiped out by poachers.

It is clear that the driving force behind rhino population declines has been large-scale commercial hunting over a long period, rather than other factors such as habitat loss. Rhino biology makes them particularly vulnerable to overexploitation. They are large, slow-growing species, with low rates of population increase. Correspondingly they are able to sustain only a low level of hunting mortality (less than 5% of the population a year for a black rhino) before declining to extinction.

3. Conservation Efforts

Controlling rhino poaching can be approached from the supply end (preventing people from killing rhinos), the demand end (eliminating consumer demand for rhino horn products), or at some point along the supply chain (blocking off trade routes). The CITES international trade ban of more than 20 years' standing has not been successful in blocking off trade routes, even if it has reduced overall volumes; as one country has clamped down, trade has shifted to others (Leader-Williams, 1992). A medicine as valued as rhino horn, used to treat serious illness, will continue to have a market even at very high prices. Thus demand is not easily reduced, despite efforts to find substitutes such as buffalo or saiga antelope horns. Whether or not demand for horn can be reduced in the long term, rhinos are so threatened currently that even a low level of exploitation in the short term could be disastrous. Thus conservation efforts have tended to concentrate on preventing the illegal killing of rhinos.

Experience of the disastrous decline in black rhino numbers over the past 20 to 30 years has shown that the key requirement for rhino conservation is small, heavily guarded reserves—an approach known as "fortress rhino." However, there has been a recent shift in perspective, particularly in the southern African countries, caused by recovery of the southern white rhino populations and the general interest in sustainable use. The costs of heavy law enforcement to keep people out of reserves are very high, financially and socially; the shoot-to-kill policies instituted by several countries for anyone found poaching in a reserve are neither popular with local people nor particularly effective (Leader-Williams and Milner-Gulland, 1993). Since rhino poaching is done mostly by commercial poachers for the international trade, not by local people hunting for meat, the usual community initiatives to provide a sustainable income for would-be poachers are not possible; and although local attitudes can help in reducing poaching (for example, if local people do not shelter poachers), communities should not be expected to carry the burden of the expensive and dangerous law enforcement needed to prevent commercial poaching.

Rhino horns could be significant sources of income to support law enforcement; rhinos do not have to be killed for their horn to be used. Pioneering dehorning programmes (e.g., in Namibia) are seen by some as the answer to the problem of rhino conservation—not only does dehorning protect rhinos from poaching by making them unattractive targets, but selling the horn could provide revenue for conservation and development programs. Thus southern African countries are pushing for a limited legal horn trade, similar to that agreed by the 1998 CITES Conference of the Parties for ivory. However, although dehorning is a potentially promising conservation tool, the issues involved are complex and not fully understood. There are worries about the effects on rhino behavior and possible increased mortality caused directly or indirectly by dehorning. Dehorning is expensive and needs to be repeated often enough to make the rhinos unattractive targets for poachers; this may need to be every 1 to 2 years. The Hwange rhinos that were dispatched so rapidly by poachers in 1991 had been dehorned 2 or 3 years previously; it may be that the horns were large enough to be attractive to poachers once again. Although the legalization of the international rhino horn trade may be extremely positive for Southern African countries, it might have negative effects on the already precarious Asian species by sending confused messages to traders and consumers.

There has been debate about the relative merits of *in situ* and *ex situ* conservation of rhinos, particularly the highly endangered Javan and Sumatran species. Rabinowitz (1995) suggests that captive breeding is not a suitable conservation option for the Sumatran rhino and Leader-Williams (1993) suggests that *in situ* conservation (protection from poachers) is a much more cost-effective method of reducing rhino mortality than captive breeding. The problem is that these species have small population sizes, causing vulnerability to stochastic extinctions, but a dearth of the high-quality population estimates needed to calculate a suitable number for removal to captivity. There are also uncertainties concerning the sustainability of captive herds, given that some species have not bred successfully in captivity, and about the ability of *in situ* conservation to protect scattered individuals from a poaching threat.

Rhinos are severely threatened mammals, for which any continuation of exploitation could lead to the extinction of one or more species. They require intensive and expensive protection to guard against this threat, on an ongoing basis. It seems highly unlikely that any truly wild rhino populations will remain extant for long into the 21st century, with zoos and heavily guarded reserves being the only places where they are likely to survive.

C. Saiga Antelopes

Like rhinos, saigas are chiefly threatened by hunting for international trade, as well as for meat. However, there are major biological differences between them, and this means that their reaction to heavy hunting and thus the emphasis of conservation efforts are also very different. Conservation efforts for saigas focused in Soviet times on promoting managed commercial harvesting; recently these efforts have collapsed, and the situation is now being reexamined.

1. Background

The saiga antelope (*Saiga tatarica*, Fig. 4) is a nomadic herding species found in the semiarid rangelands of Kazakhstan, Russia, and Mongolia (Fig. 5). It has two subspecies, *S.t. tatarica* and *S.t.mongolica*; the latter is found only in Mongolia. The Kazakhstan populations currently make up more than 80% of the species; their ecology and management are reviewed in Bekenov *et al.* (1998).

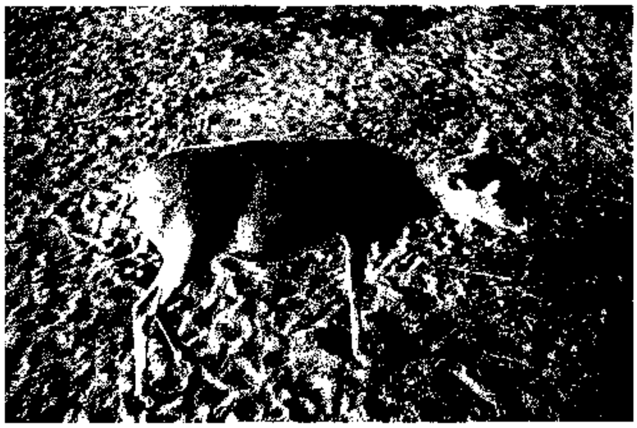


FIGURE 4 Picture of a saiga antelope. The saiga is about the size of a domestic goat, with a sandy-colored upper body and a creamy underside. This picture shows an adult female. The males have horns, which are an unusual translucent amber color. The species' most striking feature is a protuberant nose, which swells further in rutting males.

The saiga has been hunted for its meat, horns, and hide since prehistoric times. In the 18th and 19th centuries, it was hunted in large numbers by the St. Petersburg Imperial court. By the early 20th century, hunting had reduced it to near extinction. Horn prices were very high, with horns exported for use in Chinese medicine. During the Soviet period, up to 1990, the population was well managed, with legal protection and regulated commercial hunting, and grew to relative stability. However, the situation changed dramatically with the breakup of the Soviet Union. Reports on the horn trade over the past few years are now mirroring those of the late 19th century and early 20th century in noting high prices, large quantities exported to China, and worries about the effects on populations. International concern about the species led to it being listed on Appendix II (monitored trade) of CITES in 1994.

The key features of the saiga's life history that make it resilient to hunting pressure are its nomadic behavior, its ability to give birth in the first year of life and to twin consistently, and its harem breeding system. The nomadic habit of the species is a response to the harsh climatic conditions of the continental ecosystem in which it lives, allowing herds to move rapidly away from areas where there is bad weather or where other threats exist. Saigas live in desert areas in winter, where there is lower snow cover, migrating up to 600 km to the steppe areas where they spend the summer; here higher rainfall levels provide abundant grass. The females' ability to twin consistently is unusual among ungulates and means that the population can increase

FIGURE 5 Map of the range area of the saiga antelope. Adapted from Bannikov, A. G., Zhirnov, L. V., Lebedeva, L. C., and Fandeev, A. A. (1961). *The Biology of the Saiga*. Nauka, Moscow, USSR. There are four distinct populations of *S.t. tatarica*, shown as stippled areas. Three of these are in Kazakhstan, the fourth in the Autonomous Republic of Kalmykia, Russia. The other subspecies, *S.t. mongolica*, has only a few hundred individuals and is found in Mongolia.

very rapidly in good years, allowing populations to recover quickly from harsh winters or disease outbreaks. The harem breeding system, in which each adult male controls and mates with a group of 12 to 30 adult females, is important because there is an incentive for hunters to target adult males. Adult males bear horns, which are highly prized in Chinese medicine. Although heavy hunting for adult males has caused the population sex ratio to drop to less than 10% adult males (compared to 25–30% in an unhunted population), this has not led to any discernible decrease in population fecundity, possibly because the change in sex ratio is buffered by the harem breeding system (Milner-Gulland *et al.*, 1995). However, if hunting of adult males continued to depress the sex ratio, sudden collapses in fecundity might result, with serious consequences for the population. Overall, though, the saiga antelope is a resilient species that is capable of withstanding relatively heavy hunting pressure and of recovering quickly from episodes of high mortality.

2. Threats

The saiga population is still numerous, at least in Kazakhstan. However, it is under increasingly heavy pressure and concerns are growing that it may rapidly become threatened. Threats to the saiga stem from the large-scale political and social upheaval that has happened in the former Soviet Union in the 1990s, leading to a collapse in the rural economy, with dramatic reductions in livestock numbers and crumbling infrastructure. This is good news for the saiga in that it is likely to have led to a recovery of overgrazed vegetation and a reduction in contact with livestock, reducing parasite and disease transmission rates. However, this benefit is more than offset by an increase in hunting of saigas for meat, as people who have lost their livelihoods turn to saigas for food and income. As the economic crisis has also hit scientific research, population counts are no longer as frequent and reliable as they once were, so that although there are anecdotal reports of large-scale population declines in recent years, these are unconfirmed. There is also an increased risk that mass mortality from disease may strike; collapses in veterinary services in Kazakhstan have led to outbreaks of foot and mouth disease in livestock, which could cause a major epidemic in saigas. In 1967, before livestock vaccination started, the disease spread over an area of more than 100,000 km^2 in central Kazakhstan, resulting in the deaths of 50,000 saiga calves (Bekenov *et al.*, 1998).

Thus although saigas face threats from habitat loss and degradation, particularly in Kalmykia, the major current threat is illegal hunting for horns and meat, combined with the threat of mass mortality from disease. Threat levels vary; one population in Kazakhstan is under intense pressure, another is relatively unaffected because of its remoteness. In Mongolia, the threats are less severe, but are magnified because of the very small population size, making it vulnerable even to low levels of human pressure. The threat to the Mongolian population should also be taken more seriously because of its genetic distinctiveness. Thus although overall the saiga is not in imminent danger of extinction, it is under threat in most of its range. It is an important species, being the only wild ungulate present in significant numbers throughout most of its range, as well as being economically significant; this suggests that sustainable management could have significant benefits for local people.

3. Conservation Efforts

The traditional approach to mammal conservation of establishing protected areas is unfeasible for saigas. Because the saiga is nomadic, it is reliant on huge areas of relatively natural vegetation continuing to exist in the semiarid rangelands of Central Asia. Another approach that is often used is to establish protected areas in key locations, for example, in calving or mating areas. Again, this is problematic for the saiga; although they do tend to visit particular areas, they are not predictable in their use of any one area from year to year. One of the best aspects of Soviet conservation measures was the ability to establish temporary reserves (*zakazniks*) at particularly sensitive times such as the birth season, placed wherever the saiga happened to be, in which there were specific prohibitions on activities such as grazing, hunting, and driving motor vehicles. These flexible restrictions are much more likely to be of practical benefit to the saiga than the necessarily small permanent reserves that might suit other species.

However, any system of protection is only as effective as its enforcement. This relies not only on high levels of investment in law enforcement agencies, but also on the engagement of local people in conservation efforts. The importance of local people having a stake in conservation, usually through revenues from the use of these species, is well accepted nowadays. The saiga appears to be an ideal species for sustainable use, being resilient to harvesting and of high value. However, the problem in giving local people a stake in its conservation is its nomadic behavior; the range area of each saiga population in Kazakhstan is hundreds of thousands of km^2. Without a direct and predictable link between a community's conservation actions and the status of the population they are conserving, community-based conservation will not work.

This discussion suggests that saiga conservation needs to be carried out on a larger scale than the local community. Again, the previous practice of Soviet saiga managers is a good model to follow. Commercial hunting organisations were responsible for saiga management from the protection of birth areas to the processing and distribution of meat. This chain of responsibility within a single organization meant that the correct incentives existed for conservation of the resource on which future profits depended. However, problems have increased recently; large-scale selective poaching has led to the commercial organizations being unable to harvest males and thus they cannot realize the potential profits from horn sales. Underfunding of law enforcement efforts, no investment of poaching fines into saiga conservation, and a lack of support for hunting inspectors' powers of arrest have all contributed to the problem (Bekenov *et al.*, 1998). Other problems include corruption, particularly at customs, the lack of international control on saiga horn sales, and the huge areas that must be covered to protect saigas. A system that could have continued to work well for saiga conservation has not been able to cope under the stresses of sociopolitical change, leaving the saiga effectively unprotected.

The key issue identified here is the overwhelming effect that sociopolitical changes can have on efforts to conserve mammals. General economic conditions, particularly in the rural economy, or other external events such as war (as seen recently for the wildlife of Rwanda), can render even the best conservation system powerless to resist the overexploitation of a species.

D. Themes of the Case Studies

In the case studies presented here, we discussed the contrasting cases of three mammal species. In all three cases, threatening factors are modulated by the species' biology—two mammal species facing similar threats will often react differently. These differences are reflected in the conservation efforts that have been directed at each of the species.

African wild dogs were perceived as vermin and severely persecuted in the past, both inside and outside protected areas. However, although this persecution has caused the population to decline to very low levels, it is not the whole story, because other persecuted species like hyenas have not suffered the same fate. Because the wild dog lives at very low densities, it requires very large areas to maintain a viable population. Ranging widely, the species is also more vulnerable to mortality from road traffic, disease, and persecution at the edge of protected areas, so that the smaller the protected area, the higher the mortality rate from these factors.

Rhinos are also highly endangered species, but in this case, their threatened status can be firmly linked to high levels of hunting for their horns throughout recorded history. Again their biology is a contributory factor in their susceptibility to hunting, because they have very low population growth rates so cannot sustain high levels of hunting. As their populations decline, even low levels of exploitation become disastrous. Because of the nature of the threat they face, protection measures center around very intensive (and expensive) protection of small reserves where they can be guarded from poachers. This is in stark contrast to the wild dogs, which require very large protected areas.

The theme of the interaction between a species' biology and its threatened status continues with the saiga antelope. This species was hunted to very low numbers at the turn of the century to supply the market for Chinese traditional medicine, just like rhinos. However, unlike rhinos the saiga is a biologically resilient species, and saiga numbers bounced back to peak at about 1 million animals in the 1970s. Since then, heavy hunting and habitat alteration have reduced numbers, but it is only in the past few years, since the fall of the Soviet Union, that hunting threatens the species again. The saiga, like the wild dog, is wide ranging and reliant on extremely large areas of relatively undisturbed habitat for its survival; this is available for saigas at the moment, but is always vulnerable to conversion for human needs.

Critical examination of the conservation efforts directed at these species throws some light on current debates in conservation. For example, conservation efforts are currently often focused on promoting sustainable use by local communities, under the philosophy that only if a species is given a monetary value will people have the incentive to conserve it. However, none of the species described here appears a particularly suitable candidate for community-based conservation based on sustainable use. The wild dog has little obvious commercial value, the rhino has too much and thus needs expensive protection, while the migratory saiga is not easy to assign to the ownership or control of any particular community.

Another suite of conservation efforts involves intensive intervention, using techniques such as captive breeding and reintroductions. Rhino conservation has included captive breeding and ranching, and reintroductions are seen as one solution to the problems faced by wild dogs. Even saigas, which are still very numerous in the wild and do not yet have serious problems with habitat loss, are being proposed as candidates for ranching, captive breeding, and reintroduction. Our analysis suggests that these technological solutions of *ex situ* conservation are generally not useful; wild dogs need such large areas to receive adequate protection that hardly any suitable reintroduction sites are available, saigas are too flighty to be kept in captivity, and several of the rhino species have not bred successfully in captivity. The southern white rhino, which was brought back from near extinction through careful ranching and reintroduction programs, is the exception rather than the rule. Thus traditional conservation *in situ* is the best way forward for most of the populations of the species discussed in these particular case studies.

Conservation efforts for threatened mammals will always involve a suite of responses; which is appropriate in a given case depends on the interaction between the threats that it faces and the biology of the species. In the case of the species discussed, the efforts that may work best include for wild dogs, a landscape-level approach to habitat management, complemented by community-level conflict-minimization programs; for rhinos, intensive guarding of small protected areas; and for saigas, control of poaching followed by the potential reinstatement of commercial hunting. For other species, different approaches will be suitable. However, the key issue is money—protection from anthropogenic threats is expensive, and someone needs to pay. In the case of the saiga antelope, and possibly the rhino, the trophies from the species themselves could provide substantial revenues to offset conservation costs, but the wild dog is not able to fund itself in this way. In the end, conservation comes down to the willingness of people to pay for it, either directly or through the opportunity costs of not converting a species' habitat, rather than to the specifics of its biology.

V. CONCLUSIONS

Terrestrial mammals are often used as flagship species for conservation. The classic example of this is the panda used by the World Wide Fund for Nature, but tigers, elephants, and other charismatic species are also featured in publicity material about conservation because they are appealing to the public. Many people, particularly in wealthier countries, have very high intrinsic valuations of mammals and are extremely keen to preserve them, give them rights, and ensure the welfare of individual representatives of the species. The recent controversial proposal in New Zealand for legislation to extend some human rights to ape species provides an example of people regarding mammals as somehow different to other taxa. Other people, particularly those who live alongside wild mammals in poorer countries, are very keen to use them. Terrestrial mammals are often valuable species because of their meat, furs, or trophies, or destructive neighbors that invite persecution. At the same time, because of their biology, these species can be very susceptible to overexploitation.

Thus terrestrial mammals are high-profile species among both the general public and conservationists. This combination of attributes means that mammals often provide a battleground on which conservation battles of more general significance are fought. An example of this is the debate about the appropriate place for the listing of the African elephant on the CITES Appendices. This began in 1989 with a successful proposal to transfer the species from Appendix II (controlled international trade) to Appendix I (no international commercial trade), on the basis that the elephant was threatened with extinction from an uncontrolled, and largely illegal, international ivory trade. Although the international ivory trade could in itself explain the observed 12% per annum decline in the African elephant population over the period 1979—1986, there were other issues involved, including escalating conflicts between elephants and local people as elephant habitat declined. The debate about the ivory trade continues to rumble on, with the latest issue being CITES' agreement to a limited resumption of ivory sales from March 1999. However, the ivory trade debate was also the catalyst for a public airing and popularization of the issue of sustainable use, and the sovereign rights of nations to decide on the management of their own biodiversity. Popular articles explained the needs of local people to make a living from their wildlife if they are to have an incentive to conserve it in the face of growing pressure on natural resources. Others suggested that the sustainable use of elephants was either morally wrong, scientifically unwise given the current state of the African elephant population, or impractical because of the endemic corruption amongs officials and the very high value of the product.

Thus the conservation of terrestrial mammals is high profile and an area in which general conservation policy is debated and eventually made. The danger is that this policy may not be at all relevant to other components of biodiversity, with their own specific conservation requirements and biological peculiarities. Policies made with respect to one high-profile mammal species may not even be positive for other mammals. Flagship species have to be chosen wisely, since experience shows that species' conservation needs cannot be predicted from their body size or trophic position. For example, in the 1980s, India designated a network of reserves designed to protect tigers, under the banner of "Project Tiger." The project was expected to have the added benefit of protecting other species, notable the dhole (*Cuon alpinus*). This expectation was not met; dhole require much larger areas to persist than do tigers and have disappeared from several Project Tiger reserves (Woodroffe and Ginsberg, 1998).

Conservation efforts for terrestrial mammals are particularly noteworthy because of their impacts on biodiversity conservation in general. This is partly because mammals can be powerful forces structuring ecosystems. But it is also because they are often flagship species for conservation policy; this can be a mixed blessing for the conservation of biodiversity as a whole, as policy made with one particular group of species in mind will generally be inappropriate for others. Though this calls for caution in extrapolating the needs of terrestrial mammals to conservation policy on the broader scale, it is also important to remember that terrestrial mammals are particularly in need of conservation efforts in their own right.

See Also the Following Articles

CAPTIVE BREEDING AND REINTRODUCTION • CONSERVATION EFFORTS, CONTEMPORARY • EX SITU, IN SITU CONSERVATION • MAMMALS, BIODIVERSITY OF • MAMMALS, LATE QUATENARY, EXTINCTIONS OF • MARINE MAMMALS, EXTINCTIONS OF • PREDATORS, ECOLOGICAL ROLE OF

Bibliography

Bekenov, A. B., Grachev, Iu. A., and Milner-Gulland, E. J. (1998). The ecology and management of the saiga antelope in Kazakhstan. *Mammal Review* 28, 1–52.

Bere, R. M. (1955). The African wild dog. *Oryx* 3, 180–182.

Brooks, M. (1996). Chairman's report, African Rhino Specialist Group. *Pachyderm* 21, 3–4.

Childes, S. L. (1988). The status and distribution of *Lycaon pictus* in Zimbabwe. *Biological Conservation* 44, 301–316.

Crooks, K. R., and Soulé, M. E. (1999). Mesopredator release and avifaunal extinctions in a fragmented system. *Nature* 400, 563–566.

Foose, T. J., and van Strien, N. (1998). Conservation programmes for Sumatran and Javan rhinos in Indonesia & Malaysia. *Pachyderm* 26, 100–115.

Leader-Williams, N. (1992). *The world trade in rhino horn—A review*. Traffic International, Cambridge.

Leader-Williams, N. (1993). Theory and pragmatism in the conservation of rhinos. In *Rhinoceros Biology and Conservation: Proceedings of an International Conference* (O. Ryder, Ed.). Zoological Society of San Diego, San Diego, CA.

Leader-Williams, N., and Milner-Gulland, E. J. (1993). Policies for the enforcement of wildlife laws: The balance between detection and penalties in Luangwa Valley, Zambia. *Conservation Biology* 7, 611–617.

Mace, G. M., and Balmford, A. (in press). Patterns and processes in contemporary mammalian extinction. In *Future Priorities for the*

Conservation of Mammalian Diversity (A. Entwhistle and N. Dunstone, Eds.). Cambridge University Press, Cambridge.

Martin, E. B. (1996). The importance of Park budgets, intelligence networks and competent management for the successful conservation of the Greater One-horned Rhinoceros. *Pachyderm* **22**, 10–17.

Martin, E. B., and Martin, C. (1982). *Run Rhino Run*. Chatto & Windus, London.

Milner-Gulland, E. J., Bekenov, A. B., and Grachev, Iu. A. (1995). The real threat to the saiga antelope. *Nature*, 377, 488–489.

Rabinowitz, A. (1995). Helping a species go extinct. *Conservation Biology* 9, 482–488.

Robinson, J. G., Redford, K. H., and Bennett, E. L. (1999). Wildlife harvest in logged tropical forests. *Science* **284**, 595–596.

Selous, F. C. (1899). Elephant in Southern Africa. In *Great and Small Game of Africa*. (H. A. Bryden, Ed.). Roland Ward, London.

Selous, F. C. (1908). *African Nature Notes and Reminiscences*. Macmillan, London.

Woodroffe, R. (in press). Strategies for carnivore conservation: Lessons from contemporary extinctions. In *Carnivore Conservation* (J. L. Gittleman, R. K. Wayne, D. W. Macdonald, and S. Funk, Eds.). Cambridge University Press, Cambridge.

Woodroffe, R., and Ginsberg, J. R. (1998). Edge effects and the extinction of populations inside protected areas. *Science* **280**, 2126–2128.

Woodroffe, R., Ginsberg, J. R., and Macdonald, D. W. (1997). *The African wild dog: Status survey and conservation action plan*. IUCN, Gland, Switzerland.

MAMMALS (LATE QUATERNARY), EXTINCTIONS OF

Paul S. Martin
University of Arizona

GLOSSARY

extinction spasm or pulse A catastrophic burst of extinctions, peaking in less than a millennium.

first contact Initial human arrival on a landmass followed by human colonization.

late Quaternary extinction event Selective prehistoric extinctions, typically catastrophic, eliminating within the past 40,000 years two-thirds or more of large land mammals of America, Australia, and Madagascar and at least half the species of land birds on remote islands of the Pacific. Humans are present or suspected to be present in virtually all cases.

megafauna Large terrestrial vertebrates variously defined as >1, >10, and >44 kg adult body weight. Equivalent to 100 pounds and similar to the average weight of adult humans, a body size exceeding 44 kg is easily visualized and is adopted here.

quaternary The ice age of at least the past 1.81 million years, including the Pleistocene and the Holocene, the latter representing the past 10,000 years.

radiocarbon dating An isotopic or nuclear decay method for inferring age of organic materials. Carbon 14 is produced in the upper atmosphere by cosmic ray bombardment and oxidized to form $C^{14}O_2$. Distributed through the earth's atmosphere and oceans, a small percentage is incorporated into a variety of organic materials to decay with a half-life of 5700 years. By dating tree rings of known age, and other methods, radiocarbon determinations can be calibrated. Although routine application of radiocarbon dating is usually limited to dates of less than 40,000 years, ages up to 75,000 years have been measured.

I. INTRODUCTION: PREHISTORIC EXTINCTIONS

The late Quaternary extinction event (LQEE) is best known by the loss of large mammals (megafauna) from certain continents. Familiar faunal examples of the LQEE include woolly mammoths of the Northern Hemisphere, woolly rhinos in northern Eurasia, ground sloths and glyptodonts in the Americas, and diprotodonts and giant kangaroos in Australia. Comparisons between landmasses are revealing. The continents differ significantly in both the magnitude and in the timing of their extinctions. First Australia, then North and South America, and lastly Madagascar (the Island Continent) rapidly lost two-thirds or more of their large mammals (Fig. 1). In contrast, losses were comparatively gradual and much less severe in Eurasia and Africa. Although small animals of the continents seldom

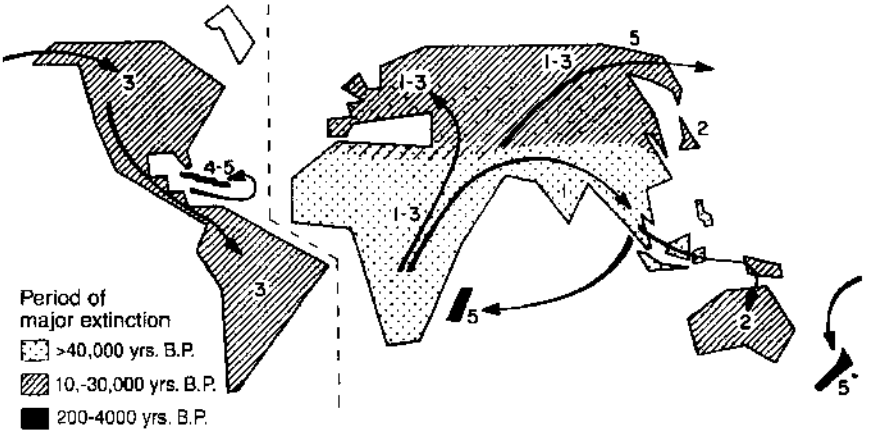


FIGURE 1 Late Quaternary extinction and prehistoric human dispersal; for numerical sequence and estimates of extinction timing, see the legend to Fig. 2 (after Martin and Steadman, 1999).

suffered extinctions, thousands of endemic mammals, birds, and reptiles vanished from oceanic islands.

At least 85 genera of large (>44 kg) mammals, half of the terrestrial megafauna of the planet, disappeared from the continents in the past 100,000 years. Now "out of sight, out of mind," the extinct mammoths and other large animals of the Quaternary coevolved with modern plant species, thereby shaping certain features of vascular plant anatomy and biochemistry. Undoubtedly, the extinct species were just as influential as living megaherbivores in determining the structure of natural communities (Owen-Smith, 1988). Before 13,000 years ago the diversity of large mammals in the New World would have resembled that currently found in African or Asian game parks. Megafaunal extinctions have made it difficult for ecologists to envision the multiple stable states prevailing in evolutionary time. In the absence of their megafauna, the savannas, grasslands, riparian habitats, and other biotic communities of the New World, Australia, and Madagascar are not functionally "natural."

Prehistoric settlement on deep-water islands in the remote Pacific began 4000 years ago. Extinctions followed, eliminating many endemic birds, reptiles, and land snails along with depletion or size reduction brought by overharvesting of near-shore fish, shellfish, and marine turtles. The relatively small number of oceanic islands that escaped prehistoric discovery suffered losses during historic time of such animals as the dodo, Steller's sea cow, giant rats, and parrots (Fig. 2). No late Quaternary extinction pulse is seen in the fossil record of the whales, other marine fauna, or in Antarctica.

An intense and often controversial search during the past 40 years for an explanation for the cause or causes of the LQEE has failed to yield any widespread consensus among paleontologists, archeologists, paleoecologists, geographers, and other interested parties, many of them convinced of the efficacy of climatic change and of climatic extinction models. In recent years, as the global pattern of the LQEE has become better known, much more attention has been focused on anthropogenic models [compare MacPhee (1999) with Martin and Klein (1984)]. Always on the scene, humans can no longer be ignored in the search for the cause or causes of "extinctions in near time."

II. THE COMPARATIVE APPROACH

In the view of most geologists the Quaternary or Pleistocene ice age of multiple glaciations embraces at least the past 1.81 million years (Ma), with the late Quaternary representing the past 200,000 years. In North America the late Quaternary includes the Rancholabrean mammalian age stage, marked by the appearance of the genus *Bison*.

With the possible exception of Australasia, most late Quaternary extinctions fall within the reach of radiocarbon assay, essentially the past 40,000 years. Fossils uncontaminated by groundwater and yielding bone collagen or gelatin are suitable for dating. Environments favorable for the preservation of bone collagen include dry caves of arid regions, the frozen ground of high latitudes, lacustrine deposits, and saline or petrolic sediments such as those at Rancho la Brea, California.

Burned bone and associated charcoal are ideal for dating, although charcoal may prove intrusive. A variety of geochemical dating methods, especially radiocarbon dating, allow biogeographers to compare both the rate of extinction within a landmass and the timing of extinctions between landmasses. A robust chronology allows intercontinental and insular comparisons of considerable value in inferring cause of the extinctions.

Although details regarding just how the extinctions were triggered remain speculative, it is increasingly clear that the two major contending explanations of what forced LQEE—that is, climatic changes and cultural impacts—can now be evaluated globally. When approached on a comparative basis, human expansion and climatic change track a punctuated extinction episode in North America and perhaps in South America, whereas climatic change is less clearly involved in Australia or Madagascar and appears to be of no significance in New Zealand and other remote Pacific islands.

A. North America

Focusing on the past 4 million years only (the Pliocene and the Quaternary), it is possible to evaluate the LQEE using age-stage divisions of biostratigraphers (Table I). The overwhelming importance of extinctions of North American large mammals at the end of the Quaternary, shown in Figs. 3 and 4 and in Table I, has been evident for at least the past 50 years.

The widespread adoption of screen washing of fossil deposits has vastly enriched the fossil record of small

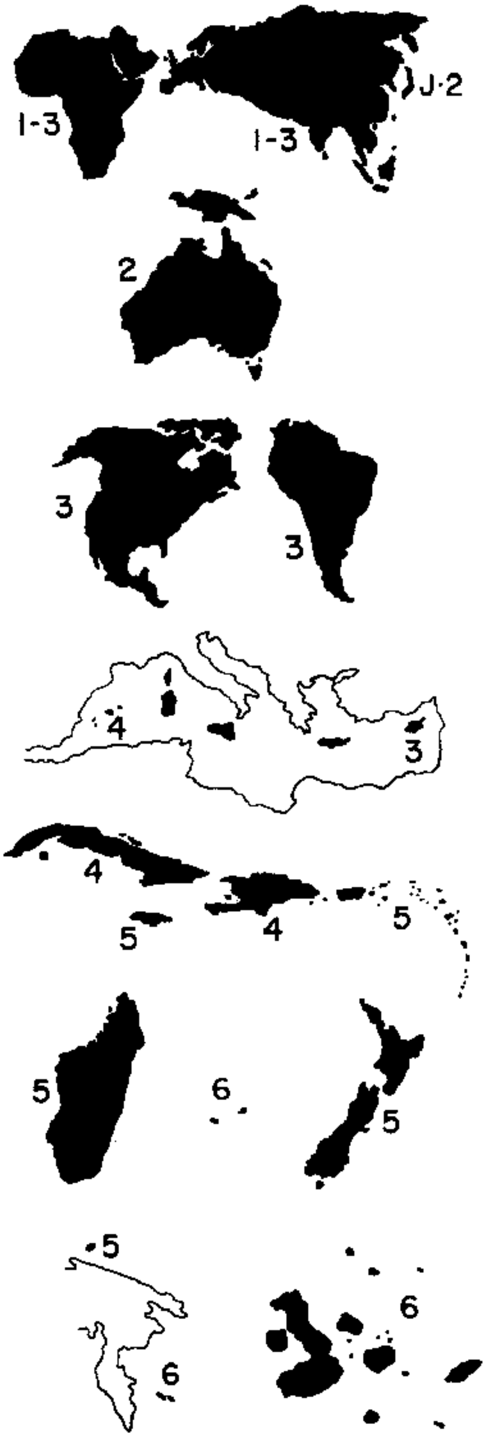


FIGURE 2 Map silhouettes of continents and islands showing time-transgressive sequence of late Quaternary extinctions and FC. *Africa and Eurasia* (1–3): Sequential extinctions of large mammals during the past 100,000 years including Nauman's elephant and giant deer in Japan [J-2] ca. 30,000 years ago. *Meganesia* (2): Humans arrive and major extinctions occur 30,000–40,000 yr bp; neither event is well dated or constrained. *Americas* (3): Well-dated evidence for human arrival and for megafaunal extinctions center on 11,000 yr bp (13,000 calendar years). *Mediterranean islands*: Epipaleothic arrival and dwarf hippo and dwarf elephant extinction on Cyprus (3) 10,500 yr bp; mid-Holocene colonization and extinction of goat antelope (*Myotragus*) on the Balearic Islands (4); FC and LQEE chronology elsewhere are uncertain. *Antilles*: Humans arrive in Cuba and Hispaniola in early or mid-Holocene (4) and in Jamaica and the Lesser Antilles in the late Holocene (5); few radiocarbon dates on extinctions. *Madagascar* (5): Humans arrive 2000 yr bp; major episode of extinction terminates 1500 AD. *Mascarenes*, east of Madagascar (6): Humans arrive 1600 AD, followed by extinctions of dodo, solitaire, other birds, and giant tortoises. *New Zealand* (5): Humans colonize by 1200 AD; the Polynesian rat perhaps colonized much earlier. Giant flightless birds (moas) are extinct by 1500 AD. *Wrangel Island* (5): Last woolly mammoths dated 4000 yr bp, 1000 years older than oldest cultural material. *Commander Islands* (6): Humans arrive 1741 AD; Steller's sea cow extinct within 30 years. *Galapagos Islands* (6): Bishop of Panama arrives 1535 AD; extinction rates increase over background by two orders of magnitude (after Martin and Steadman, 1999).

TABLE I

Large (>44 kg) Plio-Pleistocene Terrestrial Mammals of North America North of Mexico[a]

	BLANCAN				IRV			RLB		H
	1,2	3	4	5	E	M	L	E	L	
Stage duration (Ma)	1.0	0.5	0.5	0.2	0.9	0.4	0.2	0.2	0.1	0.01
Xenarthra										
+*Glyptotherium*, glyptodont			X	X	X	X	X	X	X	
+*Holmesina*, northern pampathere			X	X	X	X	X	X	X	
+*Pachyarmatherium*, ground sloth							X			
+*Eremotherium*, giant ground sloth					X	X	X	X	X	
+*Nothrotheriops*, Shasta ground sloth					X	X	X	X	X	
+*Megalonyx*, Jefferson's ground sloh	X	X	X	X	X	X	X	X	X	
+*Paramylodon*, big-tongued ground sloth		X	X	X	X	X	X	X	X	
Carnivora										
+*Borophagus* plundering dog	X	X	X	X	X					
**Canis*, dire wolf, gray wolf				X	X	X	X	X	X	X
+*Protocyon* Troxell's dog					X					
Ursus, bears		X	X	X	X	X	X	X	X	X
[+]*Tremarctos*, Florida cave bear		X	X	X	X	X	X	X	X	
+*Arctodus*, short-faced bear						X	X	X	X	
+*Chasmaporthetes*, hunting hyena	X	X	X	X	X					
+*Megantereon*, western dirktooth	X	X	X	X						
+*Smilodon*, sabertooth					X	X	X	X	X	
+*Ischyrosmilus*, Idaho sabertooth	X									
+*Homotherium*, scimitar cat					X	X	X	X	X	
+*Dinofelis*, false sabertooth		X	X	X						
[*]*Panthera*, American lion, jaguar						X	X	X	X	X
+*Miracinonyx*, American cheetah				X	X	X	X	X	X	
Felis, cougar, puma					X	X	X	X	X	X
Rodentia										
+*Procastoroides*, large beaver	X	X	X							
+*Castoroides*, giant beaver						X	X	X	X	
+*Neochoerus*, giant capybara					X	X	X	X	X	
[+]*Hydrochoerus*, Holmes's capybara							X	X	X	
Proboscidea										
+*Mammut*, American mastodon	X	X	X	X	X	X	X	X	X	
+*Stegomastodon*, stegomastodo		X	X	X	X					
+*Rhyncotherium*, rhynchothere	X	X	X	X						
+*Cuvieronius*, gomphothere			X	X	X	X	X	X	X	
+*Mammuthus*, extinct mammoths					X	X	X	X	X	
Sirenia										
+*Hydrodamalis*, Stellar's sea cow	X	X	X	X	X	X	X	X	X	X
Trichechus, manatee	X	X	X	X	X	X	X	X	X	X
Perissodactyla										
+*Cormohipparion*, extinct equid	X	X								
+*Nannipus*, gazelle–horse	X	X	X	X						
+*Plesippus*, extinct equid	X	X	X	X						
[+]*Equus*, horse species		X	X	X	X	X	X	X	X	
[+]*Tapirus*, tapir species					X	X	X	X	X	

continues

Continued

	BLANCAN				IRV			RLB		H
	1,2	3	4	5	E	M	L	E	L	
Stage duration (Ma)	1.0	0.5	0.5	0.2	0.9	0.4	0.2	0.2	0.1	0.01
Artiodactyla										
+*Megalotylopus*, large camelid	X	X	X							
+*Blancocamelus*, giraffe—camels				X						
+*Titanotylopus*, gaint camelid	X	X	X	X	X					
+*Camelops*, camel species				X	X	X	X	X	X	
+*Hemiauchenia*, llama species	X	X	X	X	X	X	X	X	X	
+*Palaeolama*, stout-legged llama							X	X	X	
+*Mylohyus*, long-nosed peccari			X	X	X	X	X	X	X	
+*Platygonus*, flat-headed peccari	X	X	X	X	X	X	X	X	X	
+*Bretzia*, false elk		X								
Odocoileus, deer species	X	X	X	X	X	X	X	X	X	X
+*Torontoceros*, extinct large cervid									X	
+*Navahoceros*, mountain deer									X	
Rangifer, caribou				X	X	X	X	X	X	X
**Alces*, moose, broad-fronted moose									X	X
+*Cervalces*, stag–moose									X	
Cervus, wapiti (elk)				X	X	X	X	X	X	X
+Tetrameryx, four-horned pronghorns					X	X	X	X	X	
+*Hayoceros*, Hay's pronghorn							X			
+*Stockoceros*, four-horned pronghorns							X	X	X	
Antilocapra, pronghorn									X	X
[+]*Saiga*, saiga									X	
**Oreamnos*, mountain goats					X	X	X	X	X	X
Ovis, bighorn or mountain sheep						X	X	X	X	X
+*Euceratherium*, shrub ox					X	X	X	X	X	
+*Soergelia*, Soergel's ox					X	X	X			
+*Bootherium*, bonnet-headed musk ox							X	X	X	
+*Praeovibos*, extinct musk ox							X			
Ovibos, musk ox									X	X
[*]*Bison*, bison species								X	X	X
+*Platycerabos*, flat-horned ox								X		
Primates										
Homo, *modern H. sapiens*									X	X
Originations	—	7	4	6	13	4	7	2	8	0
Extinctions	1	2	2	6	5	0	4	1	33	0
Total genera	18	24	26	30	37	36	43	41	48	15

[a] After Kurten and Anderson (1980), Martin and Steadman (1999), and Anderson (personnel communication). [a] Abbreviations used: IRV, Irvingtonian; RLB, Rancholabrean; H, Holocene; E, early; M, middle; L, late; +, extinct genus; *, extinct species; brackets indicate generic survival on other continents. Living genera including extinct taxa of large size that vanished with late RLB extinction include *Dasypus belli* (giant armadillo), *Canis dirus* (dire wolf), *Panthera leo atrox* (American lion), *Alces latifrons* (broad-fronted moose), and *Oreanos harringtoni* (Harrington's mountain goat). Age estimates in millions of years (Ma): Blancan 1,4.0–3.5 Ma; Blancan 2,3.5–3.0 Ma; Blancan 3. 3.0–2.5 Ma; Blancan 4, 2.5–2.0 Ma; Blancan 5, 2.0–1.8 Ma (Blancan–Irvingtonian Boundary); early Irvingtonian, 1.8–0.9 Ma; middle Irvingtonian, 0.9–0.5 Ma; late Irvingtonian, 0.5–0.3 Ma; early Rancholabrean 0.3–0.1 Ma; late Rancholabrean, 0.1–0.01 Ma; Holocene, last 10,000 years. There may not be specimen records for all cells within the temporal range of a genus as plotted here. Checklist and age estimates are courtesy of Elaine Anderson (1996), Denver Museum of Natural History, Denver, Colorado, revised from Kurtén and Anderson (1980).

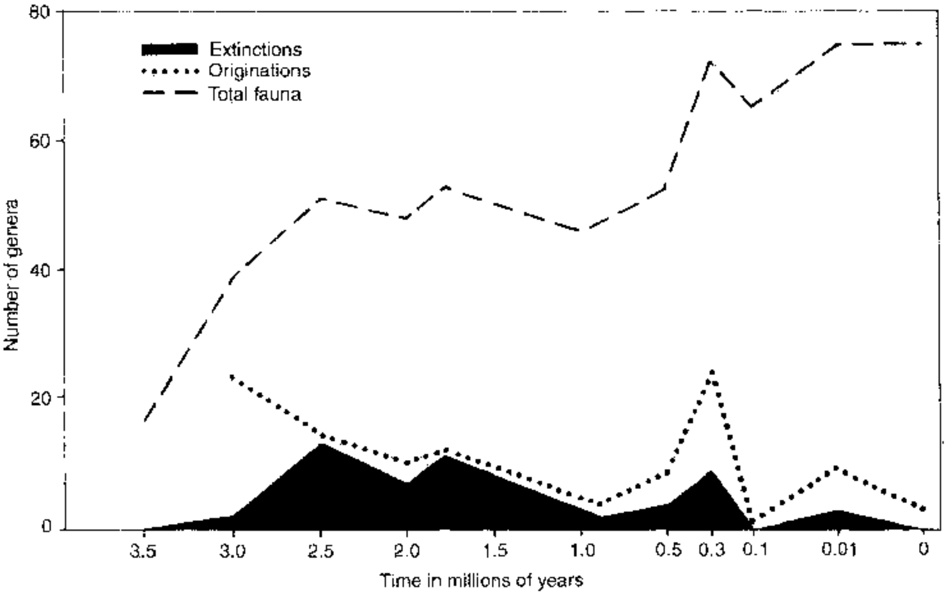


FIGURE 3 North American small mammal originations (dotted line), extinctions (solid line), and total fauna (dashed line) since the Miocene from data shown in Table I. Note the change of scale in the past 0.5 Ma (reproduced with permission from Martin and Steadman, 1999).

vertebrates. The harvest of identified small mammals is vital in supporting the conclusion that no appreciable extinctions of small mammals occurred (Fig. 3)—certainly not matching the extinction spasm of large mammals approximately 11,000 radiocarbon years ago (Fig. 4).

The fossil record is based on a much larger sample of late Pleistocene than early Pleistocene and Pliocene

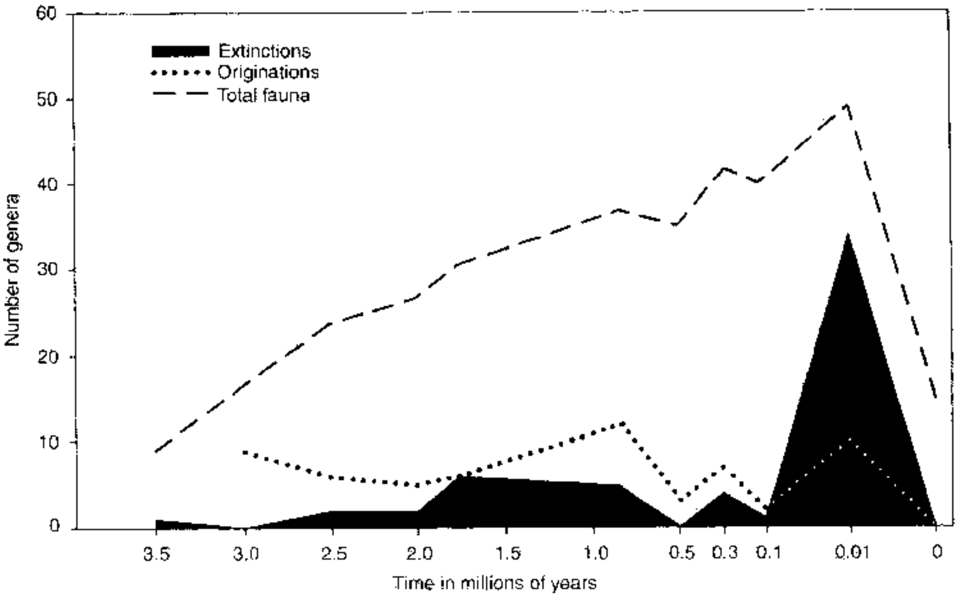


FIGURE 4 North American large mammal originations (dotted line), extinctions (solid line), and total fauna (dashed line) since the Miocene from data shown in Table I. Note the change of scale in the past 0.5 Ma (reproduced with permission from Martin and Steadman, 1999).

faunas (Table I and Figs. 3 and 4). Presumably this accounts at least in part for the increase in number of genera of both large and small mammals in the later part of the record. Although *Mammuthus*, *Equus*, and extinct *Bison* included multiple species, most of the large genera listed in Table I are monotypic (represented by one species only). Although a few survived on other continents, most large genera totally vanished (Table I).

In North America north of Mexico, 33 genera of large mammals disappeared in the late Quaternary, which constitutes two-thirds of the total late Rancholabrean fauna of 48 genera and more than all the generic extinctions recorded during the previous 4 million years prior to LQEE (Table I). Events that might trigger regional or global extinctions in the early part of the past 4 million years include intercontinental suturing of the Americas with faunal intermingling, extraterrestrial accidents such as the impact of the Eltanan bolide, and the initiation of continental glaciation. However, only at the end of the Quaternary is there a major extinction spasm—one that impoverished America's large mammalian faunas (Fig. 4).

Unlike the case of the large mammals, a much larger number of small genera (48) vanished in the past 4 million years. Within the minifauna (rat, mouse, or shrew-size mammals), which make up the majority of the 75 living genera of small mammals, no generic extinctions are known in the late Quaternary. Among the medium-sized mammals, three genera were lost: the antilocaprid *Stockoceros*, the skunk *Brachypotoma*, and the rabbit *Aztlanolagus*. The difference deserves close inspection. Although small mammals escaped virtually unscathed, the LQEE (late Rancholabrean losses) blighted all terrestrial orders of large land mammals and eliminated one, the Proboscidea (Table I).

Since the vast majority of Cenozoic (last 65 Ma) extinctions occurred long before radiocarbon time or FC, it is essential to know something of the Cenozoic pattern. Do earlier bursts of extinction match the LQEE pattern at the very end of the Cenozoic? As a result of a relatively rich fossil record of the Cenozoic, the question is readily approached in North America.

Paleontologist John Alroy (cited in MacPhee, 1999) recently analyzed 65 Ma of change in mammalian faunas through 1.0 Ma sampling bins. His data consist of extinction time rate series that are computed from a multivariate ordination of 4015 faunal lists from 2415 publications that span the Late Cretaceous through Sangamonian (last interglacial). The lists can be viewed on the World Wide Web at the North American Mammalian Paleofaunal Database: *http://homebrew.si.edu/nampfd.html*.

According to Alroy (as quoted in MacPhee, 1999, p. 120), standing diversity is defined by

> the number of species that cross each boundary between sampling bins, e.g., the time planes at 65.0, 64.0 ... 1.0 Ma. Extinctions are computed by counting the number of species in each cohort that fail to survive until the next time plane. Note that this excludes single-interval species (those that appear and disappear in the same bin), which makes the analysis less vulnerable to sampling artifacts.

Alroy's results indicate that within the past 55 Ma there were more extinctions of large species in the LQEE than there were during any earlier time. The difference in the mean body masses of the victims and the survivors peaks at the end of the Pleistocene (Fig. 5).

Although there is a general increase in mean mass of mammals through the Cenozoic, the mean mass of victims (extinct species) is highest in the past 1 million year bin within the Cenozoic (Fig. 6), which includes the LQEE. In other words, when compared with 64 turnovers earlier in the Cenozoic, the 65th has a unique property—excessive extinction of large mammals.

With the exception of bison (*Bison*) and deer (*Odocoileus*), the common fossils of large herbivores found in Quaternary deposits before the LQEE are not the

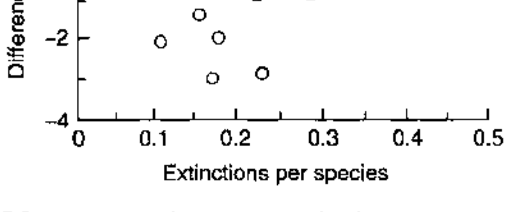


FIGURE 5 Diversity and extinction in the Cenozoic. Data are prepared by multivariate ordination and randomized subsampling. The figure plots the correlation among 1.0-Ma intervals between extinction intensity and the difference in the mean body masses of the victims and the survivors. The end Pleistocene value (●) is among the highest during the past 55 Ma (reproduced with permission from MacPhee, 1999).

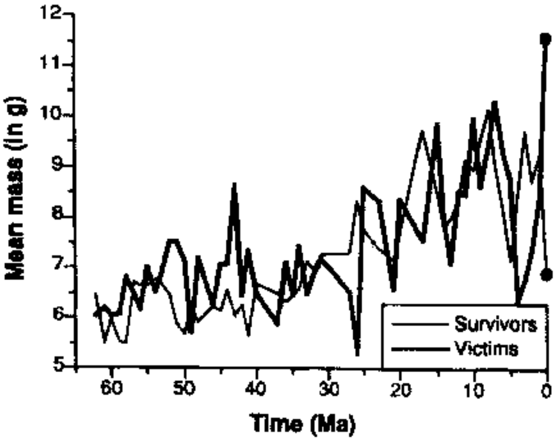


FIGURE 6 Trend through the Cenozoic in the mean body mass of victims and survivors in each 1.0-Ma interval (see legend to Fig. 5) (reproduced with permission from MacPhee, 1999).

bones or teeth of species that survived the extinctions, such as *Alces*, *Antilocapra*, *Cervus*, and *Rangifer*, but rather those that did not, especially *Camelops*, *Cervalces*, *Equus*, *Mammut*, *Mammuthus*, and *Platygonus* (Graham and Lundelius, 1994). All have terminal radiocarbon records at approximately 11,000 years ago (Stuart, 1991). One might imagine that if the LQEE were a natural catastrophe the ranges of the Holocene survivors (15 species listed in Table I) would have expanded rapidly as niche space opened following extinctions. For example, if some lethal climate condition at the end of the last glaciation eliminated their North American relatives, one might expect that afterward early Holocene warming would see the surviving populations of bison, elk, and moose expanding or at least maintaining their range. In fact, bison range shrank away from both coasts (Graham and Lundelius, 1994). Only in recent years, with the help of local introductions by game departments, has the local range and numbers of moose, elk, and black bear expanded.

The historic fauna of bison, deer, moose, elk, pronghorn, bears, wolves, etc. is traditionally regarded as the "natural" fauna of North America. However, the view from the past shows that they are an aberrant, extinction-pruned remnant, unrepresentative of the diversity or the ecological amplitude of native large mammals that the continent supported during the latter half of the Cenozoic. To view free-ranging bison, deer, moose, elk, pronghorn, bears, and wolves, the protected fauna of Yellowstone National Park in Wyoming or Denali National Park in Alaska, as representative of an American Serengeti is to seriously underestimate the diversity, productivity, and evolutionary potential of the continent.

B. South America, Australia, and Madagascar

The continents of South America, Australia, and Madagascar share a common property with North America—heavy extinction of large mammals late in the Quaternary. South America lost all large mammals more massive than a tapir (300 kg); Australia lost all mammals larger than a gray kangaroo (60 kg); and, excepting the bush pig which may have been introduced, Madagascar lost all terrestrial vertebrates larger than the living *Indri*, a lemur which weighs 7–10 kg. Timing of the extinctions varies; it was earliest in Australia, later in South (and North) America, and latest in Madagascar. Although Australian small mammals (0.5–2.0 kg) suffered historic losses, there is no indication that severe late Quaternary extinction of small mammals accompanied the loss of large ones on the continents.

1. South America

The Quaternary of South America includes the Ensenadan and Lujanian land mammal ages dated at 1.5 to 0.5 and 0.5 to 0.01 Ma, respectively. The preceding Uquian land mammal age from 2.5 to 1.5 Ma is of great interest because it immediately follows the intercontinental suturing of North and South America. The Uquian shows the effects of the famous Great American Faunal Interchange, a natural experiment in which two continents exchanged long-isolated terrestrial faunas across a new land bridge. The land bridge connecting the continents at the end of the Pliocene brought Northern Hemisphere carnivores, gomphotheres, artiodactyls, and perissodactyls into contact with South American endemic orders, the xenarthrans, notoungulates, and litopterns. Interestingly, millions of years before the intercontinental conjunction, megalonychid ground sloths managed to colonize both the Greater Antilles and North America.

Occasionally, biogeographers have sought to invoke intercontinental exchange as a cause of the LQEE. A detailed summary of the South American fossil faunas by Marshall and Cifelli (1990) shows that although small-mammal extinctions occurred in the Uquian, heavy megafaunal extinction did not occur until the Lujanian, approximately 2 million years after the exchange had begun and apparently coeval with the LQEE in North America.

South American extinctions of the late Quaternary were even heavier than those in North America north of Mexico, involving 50 genera of which 35 belong to extinct families. The losses were differential and involved all megaherbivores such as giant ground sloths, glyptodonts, gomphtheres, and two endemic orders, Litopterna and Notoungulata (Table II). It is not obvious that the interchange had anything to do with the LQEE.

Defensible radiocarbon dates on most of the South American fauna remain to be assembled and may be difficult to obtain given the poor preservation of bone collagen in the open sites where Lujanian faunas are often found. One source of ideal material for radiocarbon dating—dung, hair, and perishable tissue—has been obtained on mylodontid ground sloths from Cueva del Mylodon in southern Chile. Fifteen dates from 13,000 radiocarbon years before present (yr bp) to 10,600 yr bp support the extinction chronology for ground sloths in North America.

2. Australia

Flannery and Roberts (cited in MacPhee, 1999) report that near the time of first contact (FC) Meganesia (Australasia) lost approximately 28 genera and 55 species of large vertebrates (body weights exceeding 10 kg). The more important extinct genera of mammals in the Australasian (Australia and New Guinea) extinct megafauna are all marsupials. They include the extinct "lion" *Thylacoleo*; the wombats *Phascolomys, Phascolonus*, and *Ramsaya*; a Palorchestid, *Palorchestes*; four diprotodontids—*Diprotodon, Euwenia, Nototherium*, and *Zygomaturus*; a potoroid, *Propleopus*; and five giant kangaroos—*Fissiuridon, Procoptodon, Protemnodon, Sthenurus*, and *Troposodon*. In addition, there was an extinct giant flightless bird, the mihirung (*Genyornis*); a giant lizard, *Megalania*; an extinct horned turtle, *Meiolania*; and a giant python, *Wonambi*. Although older, the pattern of extinction in Australia resembles that in the Americas and Madagascar. In all three landmasses kill sites or processing sites are few, in dispute, or absent. As elsewhere, the lack of kill or butchering sites has deterred many Australian archeologists and paleontologists from invoking an overkill. Although no extinctions of Australian mammals less than 10 kg are known at FC, a sizable number of medium-size mammals are threatened or have vanished in historic time, presumably a side effect of European settlement in Australia (Flannery, 1994).

TABLE II

Large (>44 kg) Extinct Mammals of the Lujanian (Rancholabrean Equivalent) in South America[a]

- **Endemics**
  - Order Xenarthra
    - Family Dasypodidae
      - *Eutatus*
      - *Holmesina*
      - *Pampatherium*
      - *Propraopus*
    - Family Glyptodontidae
      - *Chlamydotherium*
      - *Doedicurus*
      - *Glyptodon*
      - *Hoplophorus*
      - *Lomaphorus*
      - *Neothoracophorus*
      - *Neosclerocalyptus*
      - *Panochthus*
      - *Parapanochthus*
      - *Plaxhaplous*
      - *Sclerocalyptus*
    - Family Megalonychidae
      - *Nothropus*
      - *Nothrotherium*
      - *Ocnopus*
      - *Valgipes*
    - Family Megatheriidae
      - *Eremotherium*
      - *Megatherium*
      - *Paramegatherium*
    - Family Mylodontidae
      - *Glossotherium*
      - *Lestodon*
      - *Mylodon*
      - *Scelidodon*
      - *Scelidotherium*
  - Order Perissodactyla
    - Family Equidae
      - *Equus*
      - *Hippidion*
      - *Onohippidium*
  - Order Artiodactyla
    - Family Camelidae
      - *Eulamaops*
      - *Hemiauchenia*
      - *Palaeolama*
    - Family Cervidae
      - *Agalmaceros*
      - *Charitoceros*
      - *Morenelaphus*
      - *Paraceros*
    - Family Tayassuidae
      - *Platygonus*
  - Order Litopterna
    - Family Macraucheniidae
      - *Macrauchenia*
      - *Windhausenia*
  - Order Notoungulata
    - Family Toxodontidae
      - *Mixotoxodon*
      - *Toxodon*
- **Pre-land bridge invaders**
  - Order Rodentia
    - Family Hydrochoeridae
      - *Neochoerus*
    - Family Octodontidae
      - *Dicolpomys*
- **Post-land bridge invaders**
  - Order Carnivora
    - Family Canidae
      - *Theriodictis*
    - Family Felidae
      - *Smilodon*
    - Family Ursidae
      - *Arctodus*
  - Order Proboscidae
    - Family Gomphotheriidae
      - *Cuvieronius*
      - *Haplomastodon*
      - *Notiomastodon*
      - *Stegomastodon*

[a] After Marshall and Cifelli (1990).

Despite the problem in direct radiocarbon dating on extinct fauna, the absence of extinct species from well-dated Tasmanian archeological sites up to 25,000 years in age supports the view that the Australian LQEE was over by then. The Franklin River region in Tasmania serves as an example. As in other parts of eastern Australia, Tasmania once harbored a variety of extinct species of giant marsupials. Although 21 dated late Pleistocene sites in caves or rock shelters in southwestern Tasmania ranging in age from 10,000 to at least 25,000 yr bp have yielded hundreds of thousands of animal bones, all are of living species and mainly of Bennett's wallaby, *Macropus rufogriseus* (J. Allen as cited in Kirsch and Hunt, 1997). If the extinct mihirung, giant macropods, diprotodonts, etc. were still alive 25,000 years ago or later, it is difficult to believe that they would not be represented in such rich zooarcheological material. This negative evidence supports the view that extinctions occurred earlier.

Until more fossil faunas are dated, there will be unavoidable uncertainty about when extinction occurred. In this regard, new environmental information from stable isotopes from samples of *Genyornis* eggshell is especially promising. The record to date indicates extinction of *Genyornis* at by 45,000 yr bp in several parts of its wide range and in more than one climatic province; these findings argue against extinctions driven by some climatic bottleneck (Miller *et al.*, 1999). In Australia extinctions long predate any hypothetical late glacial climatic forcing.

Finally, in terms of human origins, a comparison between Australia and America is instructive. In the New World archeologists have seen a variety of sites claimed to be 13,000–22,000 years or older fail to be verified or locally replicated, despite regular assertions of their proponents. In contrast, in the past two decades dozens of Australian sites have repeatedly yielded geochemical dates indicating dozens of sites older than Clovis in North America.

Given the smaller area of Australia compared with that of the New World, its much less productive soils and much more variable precipitation (Flannery, 1994), and the much smaller number of archeologists, geologists, paleontologists, and amateurs searching for artifacts and fossils, the abundance of sites 10,000–40,000 years old and older in Australia, compared with the half a dozen claims of pre-Clovis sites in the New World and their debatable status, is a red flag to environmentalists. Unless proposed early sites in the Americas are critically replicated by the geoarcheological community at large, as in Australia, the claims for a pre-Clovis occupation and a pre-Clovis culture, no matter how detailed or how often repeated, remain in limbo (Martin and Steadman, 1999).

3. Madagascar

Within the past two millennia, at a time when no other continents suffered appreciable extinctions of large mammals (Martin and Klein, 1984), Madagascar lost many large mammals, including 20 species of lemurs up to the size of a gorilla. All are larger than the largest living lemurs. In addition, Madagascar lost two hippos; the highly endemic anteater known as bibymalagasy (*Plesiorycteropus*), recently assigned to its own order; a large (bobcat-size) species of carnivore (*Cryptoprocta spelea*); a rabbit-size rodent (*Hypogeomys australis*); two genera of giant flightless birds; and two giant tortoises, one exceeding 100 kg in mass (MacPhee and Marx as cited in Goodman and Patterson, 1997). Although most if not all of these extinctions in Madagascar coincide with human colonization, none have been found in well-defined kill sites, leading MacPhee and Marx to model hyperdisease. They generalize their model to include all lands of prehistoric human colonization. Despite its implausibility, the model invites testing since the fossil record has yet to reveal how LQEE occurred. In the Americas the lack of many Clovis kill sites may be overcome by modeling "blitzkrieg," a rapid elimination of preferred prey, leaving minimal field evidence, by a potent human predator (Martin, 1990).

If aridification contributed to extinctions in Madagascar, it was evidently a potential problem only in the southwest. Sudden overkill by a "blitz," a sweep of the vulnerable megafauna, seems unlikely if the start-up of colonization took more than 1000 years. Robert Dewar (as cited in Goodman and Patterson, 1997) suggested that the impact of wild cattle on the native fauna rather than human hunting forced the extinctions.

All this may be resolved eventually by enrichment of the stratigraphic and geochronological record, including a major effort at extending the radiocarbon chronology of the extinct Malagasy megafauna into glacial times to determine if late glacial climatic change forced extinctions long in advance of human presence. Radiocarbon dates of 13,000 and 26,000 yr bp on a large (~70-kg) extinct lemur, *Megaladapis* (Simons as cited in Goodman and Patterson, 1997), indicate that glacial-age faunas exist. Apart from *Megaladapis*, virtually nothing is known of the glacial-age faunas of Madagascar. Like other extinct genera, *Megaladapis* survived until the end of the Holocene and may have been known historically.

Unlike Madagascar, rich glacial-age faunas are known in New Zealand, Tonga, and Hawaii. To date, no glacial-age extinctions have been detected in these islands, strengthening the case for an anthropogenic agency as the unique cause of LQEE (Martin and Steadman, 1999).

Perhaps the most interesting case of comparative extinctions is the global decline of giant tortoises. Giant tortoises can be regarded as extremely vulnerable to human foragers. They survived historically only on remote warm-water oceanic islands that escaped prehistoric colonization, such as the Mascarenes, Seychelles, and Galapagos. In the late Holocene in Madagascar, giant tortoises vanished with the giant lemurs. In Africa they disappeared much earlier, perhaps more than 2 million years ago, an event that may represent the earliest example of an anthropogenic forcing of megafaunal extinctions.

C. Eurasia and Africa

Compared with the sweeping, catastrophic loss of late Quaternary megafauna in the New World, Madagascar, and arguably Australia, the losses in Eurasia were relatively minor and they occurred gradually (Fig. 7; Martin and Steadman, 1999; Stuart, 1991). Outside the tropics, the extinct Eurasian fauna can be divided into two parts. One was a "warm" or interstadial fauna with straight-tusked elephants (*Elephas* [*Palaeoloxodon*] *antiquus*), temperate rhinoceroses (*Stenorhinus hemitoechus* and *S. kirchbergensis*), hippo (*Hippopotamus amphibius*), and a cave bear (*Ursus spelaeus*). The other was a "cold" or stadial fauna with woolly rhinoceros (*Coelodonta antiquitatis*), woolly mammoth (*Mammuthus primigenius*), and musk ox (*Ovibos moschatus*). This stepwise sequence of change in temperate and boreal faunas of Eurasia is gradual compared with the sudden extinction of both cold- (woolly mammoths and woodland musk oxen) and warm-adapted species (tapir and jaguar) in North America.

Intriguing differences in the pattern of Old and New World extinction are best illustrated in the case of proboscideans. Extinction of straight-tusked elephants did not begin until early in the last cold stage of the late Quaternary. Extinct throughout their European range by 60,000 years ago (Stuart, 1991), *Elephas* persisted in Japan until 30,000 years ago or less. The largest Holocene survivors in the megafauna of Japan were Sika deer (*Cervus*), wild boar (*Sus*), and black bear (*Selenarctos*). Although dwarf relatives of *Elephas* evidently did not survive on Mediterranean islands as late as the Holocene, as was once thought (Martin and Klein, 1984), mid-Holocene woolly mammoths have been found on Wrangel Island off Siberia, 6000 of years younger than any other known mammoths—contemporary with the Pharaohs. The Akrotiri *Aetokremnos* bone cave, an epipaleolithic site on Cyprus, yielded dwarf straight-tusked elephants along with abundant bones of dwarf hippo. The deposit is radiocarbon dated at 10,500 years old. Just as Wrangel Island served as a refugium for the last of the boreal woolly mammoths, thousands of years after they were extinct throughout the rest of their range, Japan and Cyprus apparently sheltered the last populations of temperate Eurasian *Elephas* long after its extinction on the continent (Martin and Steadman, 1999).

Unlike the slow pace of extinction of the Proboscidea in northern parts of Eurasia, in North America prior to an extinction spasm 10,000–11,000 yr bp the animals suffered no loss of taxa, no clear-cut reduction in range, and no apparent decline in numbers (see maps in Graham and Lundelius, 1994). Had American Proboscidea followed the Old World extinction pattern for elephants, the temperate elephant of midlatitudes, the Columbian mammoth, would have predeceased high-latitude extinction of woolly mammoths by tens of thousands of years while the order would endure in tropical America, represented by a living species.

Both ivory carvings and cave paintings depicted by Stone Age artists reveal their intimate knowledge of the anatomy and behavior of Old World woolly mammoths and many other animals. Although field evidence of mammoth kills is scarce, low rates of increase of modern megaherbivores means that modern elephants, rhinoceros, and hippo would have been highly vulnerable to human predation (Owen-Smith, 1988). Even with governmental protection in this century, the future of surviving megaherbivores is by no means ensured. The gradual decline of Eurasian mammoths could readily have resulted from a minimal amount of human hunting or interference, especially if younger age classes were targeted.

Although the nature of the association and its meaning in modeling insular faunal extinctions in the Mediterranean are uncertain, the record at Akrotiri *Aetokremnos* suggests a Paleolithic "commando raid" in which Stone Age hunters or foragers overran an island, slaughtered preferred prey (dwarf hippo, dwarf elephant, and endemic deer), did not find sufficient resources for a sustainable economy, and soon left with no other evidence of their passage besides the remarkably rich contents of one cave. Possibly Crete, Ireland, Wrangel Island, and the Channel Islands of California also lost megafauna to commando raids that left scant evidence or at least none discovered to date. Given the scarcity of convincing kill or processing sites on continents, the rarity of such features on offshore islands is not surprising. Only in the remote Pacific such as on the low limestone islands in Tonga are extinct animals (bones of megapodes) abundantly associated with artifacts of the first colonists. In other cases (Hawaii and Madagascar), such associations are rare or unknown (MacPhee, 1999).

Africa has been held up to modelers of overkill as a

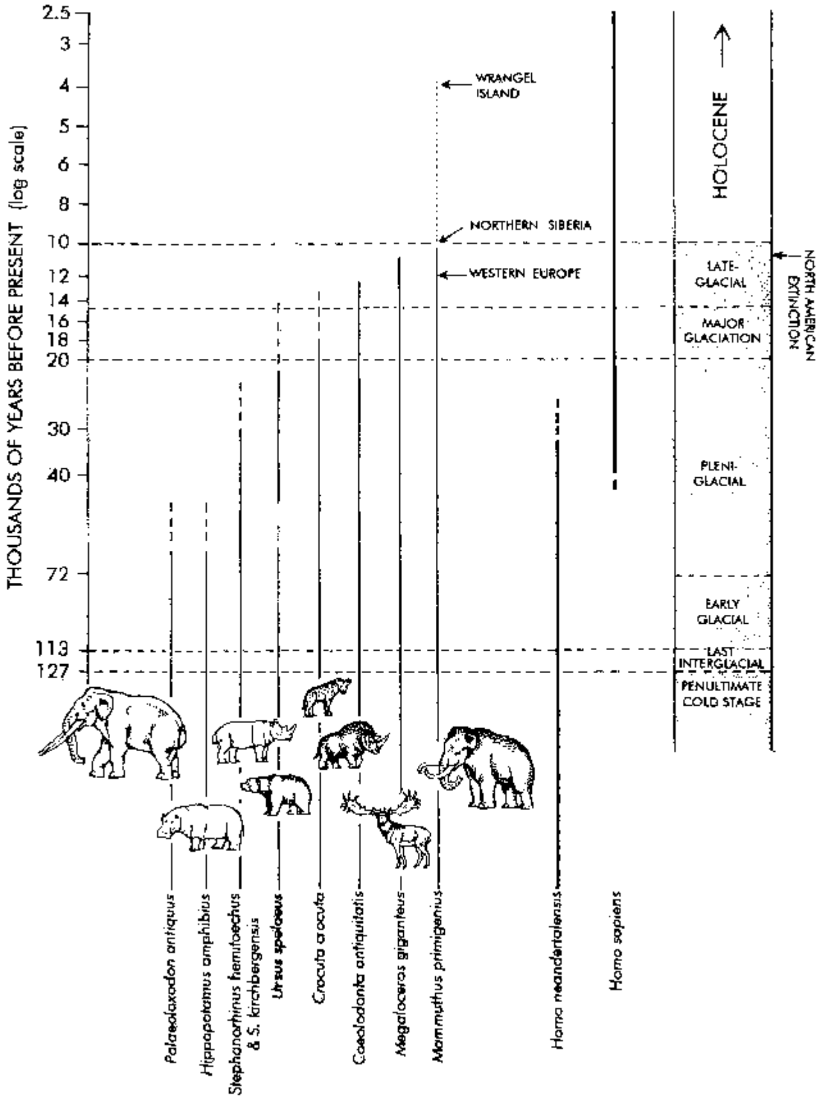


FIGURE 7 Extinction or extirpation chronology of large mammals in northern Eurasia. Note how the inferred extinction dates are very staggered compared with the strong pulse of extinctions in North America (right). During at least 50,000 yr bp the loss of Eurasian large mammals was sequential; the loss in North America was sudden. NE, northern and central Europe; It, Italy; Ja, Japan; SF, southern France; Sp, Spain; Lv, Levant; E, England; Sc, Scandinavia; Ir, Ireland; NS, north-central Siberia; *Elephas* (*Palaeoloxodon*) *antiquus*, straight-tusked elephant; *Hippopotamus amphibius*, hippopotamus; *Stenorhinus hemitoechus* and/or *S. kirchbergensis*, temperate rhinoceroses; *Ursus spelaeus*, cave bear; *Crocuta crocuta*, spotted hyena; *Coelodonta antiquitatis*, woolly rhinoceros; *Megaloceros giganteus*, giant deer; *Mammuthus primigenius*, woolly mammoth [after MacPhee (1999) and Stuart (1991)].

conundrum. If people are such effective exterminators, how did the African megafauna survive? However, the argument cuts two ways. If late Pleistocene climatic change played a major role in triggering megafaunal extinctions elsewhere, how did so many large animals manage to survive in Africa and Asia, continents no more immune to Pleistocene climatic change than other corners of the globe?

One reply to the conundrum is that in Africa and Asia there was sufficient time to evolve a balanced predator–prey relationship, perhaps with human populations locally suppressed by sleeping sickness and other endemic human diseases and on occasion by intertribal buffer zones (Martin and Klein, 1984). The extinction record in both Africa and Asia is so unlike that found elsewhere on the planet that important differences in historical biogeography and human ecology may be postulated. The record of gradual human evolution and radiation confined to the Afro-Asian landmass, ending in the late Quaternary with explosive expansion of human colonists onto all other temperate landmasses, fits the model of a time transgressive overkill.

D. Oceanic Islands

Late Quaternary losses were limited neither to continents nor to mammals. They swept through oceanic islands. For example, on the larger oceanic islands the extinction of large terrestrial vertebrates included flightless moas in New Zealand; gorilla-size extinct lemurs, extinct hippo, giant tortoises, and giant flightless birds such as *Aepyornis* in Madagascar; house cat to bear-size ground sloths in the West Indies; dwarf elephants and dwarf hippo on islands of the Mediterranean; and dwarf elephants and/or giant tortoises on oceanic islands beyond the continental shelf in southeastern Asia such as Timor, Flores, and Sulawesi. In one unusual case, prehistoric extinction only 4000 years ago eliminated elephants (woolly mammoths) not from a deep-water island but from a shelf island, Wrangel, in the Arctic Ocean (Fig. 2).

With their discovery 25 years ago on Hawaii of a flightless goose (*Thambetochen*) and a flightless ibis (*Apteribis*), ornithologists Alexander Wetmore and Storrs Olson triggered the search for unknown extinctions on Pacific archipelagos. In the following years, archeologists and paleoecologists began to uncover rich fossil faunas reflecting prehistoric (Holocene) extinctions. During the past two decades their efforts on the Cooks, the Marquesas, the Kingdom of Tonga, and the Line Islands, to name a few examples from the South Pacific, have yielded many extinct species of small birds, especially flightless rails, pigeons, and parrots (Steadman, 1995; Steadman as cited in Kirch and Hunt, 1997).

On the basis of fossil deposits from a sample of the 800 islands in the Pacific that are more than 1 km^2 in area, Steadman (1995) estimates the loss of 10 species or populations from each—a total loss of 8000 species or indigenous populations of land birds (including an estimated 2000 endemic species of taxonomically distinctive flightless rails). In addition to human harvesting and possible surplus killing, the introduction of Pacific rats may have eliminated many flightless rails and other birds. In some cases, the rats may even have colonized in advance of their human vectors (Holdaway as cited in MacPhee, 1999).

Extensive colonies of petrels, shearwaters, and other seabirds vanished from many Pacific islands, dimming the avian beacon which helped guide the prehistoric explorers to remote uninhabited islands (Steadman as cited in Kirch and Hunt, 1997). The fossil records of the relatively small number of islands that experienced FC and heavy extinction within historic time, such as the Galapagos, the Mascarenes, and remote Atlantic Islands such as St. Helena, have yet to reveal any pulse of prehistoric (Holocene) extinctions to match those found on islands that were colonized prehistorically.

New Zealand provides the best fossil record of island extinctions, starting with the loss of flightless wrens and small petrels, which were eliminated by accidental Polynesian introduction of Pacific rats (*R. exulans*), and beginning perhaps 1000 years in advance of Polynesian settlement. After Captain Cook's arrival in 1770, more species of mammalian predators reached New Zealand, including the dog, Norway rat, feral pig, feral cat, house mouse, black rat, ferret, and stoat (Holdaway as cited in MacPhee, 1999).

To bring prehistoric extinctions into focus, it is necessary to scan the globe and to probe the magnitude, timing, pattern, and natural history of all prehistoric losses on all landmasses, continents, and islands. This simple comparative approach in paleontology opens new vistas—ones that the popular MacArthur–Wilson model of island biogeography, based on modern distributions, will not reveal. High rates of vertebrate extinction on oceanic islands in the past five centuries, the focus of attention by conservation biologists (MacPhee and Flemming as cited in MacPhee, 1999), pale in magnitude to the extinction spasm of the previous four millennia.

Can extinctions of the LQEE be reversed? They certainly can if we do not insist on replacing lost species with taxonomically identical populations. On oceanic islands restoration ecologists need to consider restarting evolution of lost lineages of rails, using Guam rails on islands once occupied by flightless rails. In North America, a Pleistocene park should minimally include proboscideans, camelids, and equids as well as bison and other historic megafauna that survived the LQEE.

III. PATTERN AND CAUSE

Through an understanding of the pattern and timing of late Quaternary extinctions it may be possible to establish their cause. As the chronology, stratigraphy, paleoecology, and global pattern of the extinctions become better known, the ultimate questions of rate are clarified. For example, Darwin's view in *Origin of Species* that prehistoric extinctions of large mammals occurred gradually and sequentially is supported by the LQEE in Eurasia (Fig. 7). Although an extraterrestrial accident such as an asteroid or cometary impact is often linked to some earlier mass extinctions, especially at the end of the Cretaceous, the space rock scenario can be ruled out in the late Quaternary. Over radiocarbon time the global pattern of loss was sequential or time transgressive, from more than 30,000 years ago in Australia to 1000 years ago in New Zealand (Figs. 1 and 2), virtually eliminating any possibility of a one-shot global climatic catastrophe, such as a unique lethal cold snap, hyperdrought, or a great flood.

The prehistoric extinctions reviewed here are increasingly suspected of being the preamble of a vastly larger number to follow, a true mass extinction event in the making. Its cause would be anthropogenic.

A. Anthropogenic Models

The two major competing hypothesis or models to LQEE involve either anthropogenic or climatic forcing or both in combination (Burney, 1993; MacPhee, 1999; Martin and Klein, 1984). The anthropogenic model includes a variety of possible human impacts: direct overkill, surplus killing, a predator pit (doomed but still functioning predators, such as saber cats, reinforcing human predation), introduction of pandemic disease ("hyperdisease"), and human-initiated changes in habitat. The anthropogenic model is based on the close timing of extinctions to the global spread of anatomically and behaviorally modern people. In addition, some propose that Neanderthals or earlier hominids also had a role both in shaping the evolution of large mammal communities in the Old World and in the extinctions of the most vulnerable prey species providing rich resource packages such as giant tortoises, animals whose antipredator strategies were ineffective against humans. The kinds of animals lost in the LQEE appear to represent species known to be or likely to have been preferred prey.

A new player among the possible contenders for anthropogenically driven LQEE pulses is infectious disease, brought in at FC by human colonists themselves or, more likely, everything and anything that came along with them, such as domesticated, commensal, and synanthropic species. Especially on oceanic islands, the impact of human colonists would be amplified by the introduction of aliens, such as the Pacific rat *R. exulans*, chickens, domestic pigs, and dogs, along with their exotic diseases. The logic of hyperdisease is similar to that of classic overkill, but it contemplates humans as passive rather than active agents in causing faunal crashes. The basic idea is that "emerging" diseases, introduced into species without any natural immunity to them, would have induced incredible levels of mortality in susceptible populations, conceivably leading to their extinction. Once a "new" infectious disease process got started in a susceptible population, it would run its course quite independently of the rate or direction of human expansion.

Critics of anthropogenic models note that kill sites indicating human predation on and processing of allegedly preferred prey are rare, with few in the Americas and none in Australia or Madagascar and few in the Americas. The case of Madagascar is especially curious because its losses occurred only within the past 2000 years. Skeptics challenge the chronology of extinction, denying that within North America or Australia all extinct large mammals vanished abruptly, on the heels of human colonization, as implied by overkill. Some contend that humans arrived in both continents significantly before extinctions occurred. If human arrivals triggered megafaunal extinctions in the New World and Australia, critics ask, how do we explain the coexistence of *Homo sapiens* and large mammals in Africa and tropical parts of Eurasia? Vertebrate paleontologists have traditionally appealed to climatic or environmental changes as the main forcing function for many mammalian extinctions. During the past 65 million years the vast majority of extinctions occurred before any possibility of human involvement.

B. Climatic Models

Climatic models are based on the highly variable nature of late Quaternary environments with rapid switches from cold or cold and dry to warm and wet accompanied by changes in CO_2 pressure. In some cases, such as the Allerod–Younger Dryas shift within the late glacial, the switch appears to coincide with megafaunal extinctions, especially in North America. At least one habitat, the steppe–tundra of polar regions in the Northern Hemisphere, has been identified as an extinct biome (Guthrie as cited in Martin and Klein, 1984) whose end doomed woolly mammoths and other subarctic megafauna. Some LQEE models invoke a switch from less extreme to more extreme seasonality with out-of-step breeding cycles eliminating ruminants whose life cycles could not accommodate the climatic changes (Kilte and Graham and Lundelius as cited in Martin and Klein, 1984).

Critics discount late Quaternary climatic change as a forcing function since the paleoclimatic record of the Quaternary is rich in rapid and severe changes long before as well as during episodes of extinction. Fewer megafaunal extinctions occurred in the 3.5 million years combined, before human arrival, than in the late Quaternary, suggesting that species in Quaternary biotas were buffered against environmental switches typical of the last ice age. In any case, a late Quaternary extinction spasm is not evident in the oceans or in freshwater habitats and involves small animals only as parasites or on oceanic islands on which the reduced area would have amplified human impacts. Finally, the character of LQEE universally points to species that would either be preferred prey for human foragers and predators or be vulnerable (as in the case of minute endemic island snails) to aliens introduced from the continent.

Although details of how prehistoric humans might have triggered prehistoric extinctions are not easy to interpret from the fossil record, on a global scale the LQEE strongly reflects a deadly syncopation between extinction and human arrival. Viewed from the Cenozoic, Alroy (as quoted in MacPhee, 1999, p. 105) observed,

> The event's timing, rapidity, selectivity, and geographic pattern all make good sense according to the anthropogenic model and no sense at all otherwise. I believe that the overkill hypothesis, at least in general terms, already has been "proven" as thoroughly as any historical hypothesis can be.

See Also the Following Articles

EXTINCTION, CAUSES OF • MAMMALS, BIODIVERSITY OF • MAMMALS, PRE-QUATENARY, EXTINCTIONS OF • MASS EXTINCTIONS, CONCEPT OF • MASS EXTINCTIONS, NOTABLE EXAMPLES OF

Bibliography

Burney, D. A. (1993). Recent animal extinctions: Recipes for disaster. *Am. Sci.* 81, 530–541.

Flannery, T. F. (1994). *The Future Eaters*. Reed Books, Melbourne.

Goodman, S., and Patterson, B. (Eds.) (1997). *Natural Change and Human Impact in Madagascar*. Smithsonian Institution Press, Washington, DC.

Graham, R. W., and Lundelius, E. L. (1994). *Faunmap: A Database Documenting Late Quaternary Distributions of Mammal Species in the United States*, Scientific Papers Vol. 25, Nos. 1 and 2. Illinois State Museum, Springfield.

Kirch, P. V., and Hunt, T. L. (Eds.) (1997). *Historical Ecology in the Pacific Islands: Prehistoric Environmental and Landscape Change*. Yale Univ. Press, New Haven, CT.

Kurtén, B., and Anderson, E. (1980). *Pleistocene Mammals in North America*. Columbia Univ. Press, New York.

MacPhee, R. (Ed.) (1999). *Extinctions in Near Time*. Kluwer/Plenum, New York.

Marshall, L. G., and Cifelli, R. L. (1990). Analysis of changing diversity patterns in Cenozoic land mammal age faunas, South America. *Paleovertebrata* 19, 169–210.

Martin, P. S. (1990). 40,000 years of extinctions on the "planet of doom." *Palaeogeogr. Palaeoclimatol. Palaeoecol. (Global and Planetary Change Section)* 82, 187–201.

Martin, P. S., and Klein, R G. (Eds.) (1984). *Quaternary Extinctions: A Prehistoric Revolution*. Univ. of Arizona Press, Tucson.

Martin, P. S., and Steadman, D. W. (1999). Prehistoric extinctions on islands and continents. In *Extinctions in Near Time* (R. MacPhee, Ed.). Kluwer/Plenum, New York.

Miller, G. H., Magee, J. W., Johnson, B. J., Fogel, M. L., Spooner, N. A., McCulloch, M. T., and Ayliffe, L. K. (1999). Pleistocene extinction of *Genyornis newtoni*: Human impact on Australian megafauna. *Science* 283, 205–208.

Owen-Smith, R. N. (1988). *Megaherbivores: The Influence of Very Large Body Size on Ecology*. Cambridge Univ. Press, Cambridge, UK.

Steadman, D. W. (1995). Prehistoric extinctions of Pacific Island birds: Biodiversity meets zooarchaeology. *Science* 267, 1123–1131.

Stuart, A. J. (1991). Mammalian extinctions in the Late Pleistocene of Northern Eurasia and North America. *Biol. Rev.* 66, 453–562.

MAMMALS (PRE-QUATERNARY), EXTINCTIONS OF

W. A. Clemens
University of California, Berkeley

GLOSSARY

background extinction Rates of extinction that characterized the major part of the evolution of life.

clade A group of organisms including its common ancestor and all its descendants.

crown group A clade including a group of modern species, the common ancestor, and all its descendants, including extinct lineages.

extinction The termination of a lineage of organisms.

mass extinction A short period of Earth history in which the rate of extinction reached an exceptionally high level.

paraphyletic A group of organisms that includes the common ancestor but not all of its descendants.

EXTINCTION HAS PLAYED A major role in shaping the course of mammalian evolution from the differentiation of their early synapsid ancestors, during the interval between 354–290 million years ago, up to the Quaternary, approximately 1.8 million years ago. Throughout their history, relatively low rates of extinction contributed to a continuing background of evolutionary change. Major decreases in diversity occurred at times of mass extinctions, when many lineages were terminated during relatively short periods of time, or during periods of longer duration characterized by high levels of faunal turnover. Such major events divide mammalian evolution into three phases, which will be analyzed after a brief introduction to the history of the synapsids and the general patterns and causal factors of extinction.

I. MAMMALS AND OTHER SYNAPSIDS

The evolutionary history of mammals and their ancestors, which constitute the Synapsida, began in the Carboniferous Period approximately 354–290 million years ago (Ma) (Table I). During this interval terrestrial vertebrates underwent their first evolutionary radiation. One of these lineages, characterized by a specialized type of egg (the amniote egg), split into two clades—

Encyclopedia of Biodiversity, Volume 3

TABLE I

Geological Timescale for the Interval of Earth History Considered in the Text[a]

Era	Period	Epoch	Age of boundary (millions of years)
	Quaternary	Pleistocene	1.8
		Pliocene	5.3
		Miocene	23.8
Cenozoic	Tertiary	Oligocene	33.7
		Eocene	54.8
		Paleocene	65.0
	Cretaceous		144
Mesozoic	Jurassic		206
	Triassic		248
	Permian		290
Paleozoic	Carboniferous		354

[a] Data taken from the 1999 Geological Time Scale, Geological Society of America.

reptiles (including birds) and synapsids (including "nonmammalian synapsids" and crown group mammals; Fig. 1). Early members of these clades are distinguished by different structural patterns of the posterior parts of their skulls. Although clearly part of the ancestry of mammals, early nonmammalian synapsids lacked specialized features of skull structure that characterize members of the crown group Mammalia—for example, only one pair of bones (the dentaries) making up the lower jaw or the presence of three bones (malleus, incus, and stapes) in the middle ear. Direct evidence is not preserved in the fossil record, but it is probable that these early synapsids lacked hair, mammary glands in the females, and endothermy—traits commonly used to distinguish modern mammals from reptiles.

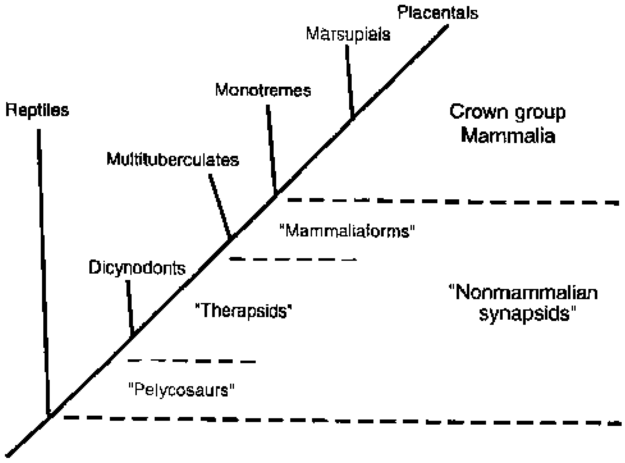


FIGURE 1 Diagrammatic representation of the phylogenetic relationships of the groups of "nonmammalian synapsids" and crown group Mammalia. Names of paraphyletic groups are enclosed in quotation marks.

Unfortunately, in many older and some recent texts, early synapsids are referred to as "mammal-like reptiles." These early synapsids definitely were not members of the ancestry of reptiles. Morphological similarities to reptiles are the product of inheritance and retention of traits found in their common ancestor. In the following discussion, these early synapsids will be dubbed "nonmammalian synapsids," which accurately expresses the fact that they were members of the early part, the stem, of the synapsid clade but not members of the crown group Mammalia (Fig. 1).

In a broad-brush summary, the evolutionary history of the synapsids can be divided into three phases. Beginning in the Carboniferous, the first evolutionary radiation of "nonmammalian synapsids" was interrupted by the mass extinction event used to mark the Permian–Triassic boundary. The surviving lineages of "nonmammalian synapsids" radiated again and regained their position as a dominant group of terrestrial vertebrates. Prior to the next mass extinction marking the Triassic–Jurassic boundary (approximately 206 Ma), almost all lineages of "nonmammalian synapsids" became extinct. The following Jurassic and Cretaceous Periods, which lasted from approximately 206 to 65 Ma, constitute the "Age of Dinosaurs." Within this interval all surviving lineages were limited to animals of small body size, but it was a time of major evolutionary radiation that included the origin of the crown group Mammalia. The Cretaceous ended with another mass extinction event; this was the time of extinction of dinosaurs other than birds. During the following Tertiary Period (65–1.8 Ma), mammals flourished and their diversity rapidly increased. Biogeographic differentiation, extinction events, and periods of faunal turnover complicated Tertiary evolutionary radiations of mammals. Changes in global climate appear to have been among the causal factors of these extinctions and faunal turnovers, which led to the origin Quaternary mammalian faunas.

II. PATTERNS AND CAUSES OF EXTINCTIONS

Although discussed in much more detail later, a few issues relating to the nature and causal factors of extinctions need to be reviewed here as background for analysis of the role of pre-Quaternary extinctions in shaping the course of mammalian evolution.

A. Extinctions and Pseudoextinctions

In any attempt to assess changes in biodiversity resulting from extinctions, care must be taken to differentiate between true, biological extinctions and what have been termed pseudoextinctions. In a biological sense, extinction is the termination of the group or groups of organisms comprising a species. No longer is there successful reproduction and survival of the young to sexual maturity, population size dwindles to zero, and the species, as well as the lineage it represents, is lost from the biota.

Pseudoextinction refers to an artifact produced by zoological nomenclature; a named unit becomes extinct but the lineage it represents survives. The taxonomy (the system of recognition and naming of groups of organisms) of the Equidae, the family that includes modern horses, asses, and zebras, provides many examples of pseudoextinction. The oldest known species of the family are a group of lineages included in the genus "*Hyracotherium*" (= "*Eohippus*"). These were diminutive, multitoed inhabitants of tropical forests. The fossil record documents increase in body size, loss of toes, and other significant morphological changes in descendants of "*Hyracotherium*"—for example, "*Orohippus*," "*Miohippus*," "*Merychippus*," and "*Dinohippus*." These extinct genera are members of the ancestry of modern horses, zebras, asses, and other closely related species, all of which are included in the genus *Equus*.

The genus "*Hyracotherium*" is composed of the common ancestor of all the species referred to "*Hyracotherium*," including the lineage leading to "*Orohippus*," the next member of the equid family tree. In the jargon of modern taxonomy, "*Hyracotherium*" is paraphyletic, a unit that includes its common ancestor and some but not all of the common ancestor's descendants; in this case, "*Orohippus*" is excluded. All genera of the family tree ancestral to Equus are now extinct. The extinctions of these paraphyletic groups, however, did not mark the termination of the equid lineage; they are products of the system of recognizing and naming organisms. (To highlight their status, the names of paraphyletic groups are enclosed in quotation marks.)

Current studies of patterns of extinction usually do not differentiate between biological (lineage) and pseudoextinctions. Therefore, the relative proportions of these two types of extinction in the fossil record are uncertain. Distinguishing between them might significantly modify some current interpretations of the patterns and processes of evolution.

It also must be stressed that in most studies of pre-Quaternary extinctions, species are not the basic units considered. Usually, these studies analyze the differences in numbers of genera or families, a procedure that opens the possibility of significantly muting patterns of change in biodiversity. For example, if a genus consists of only one species, loss of the species results in the loss of the genus. On the other hand, if the genus contains many species, the loss of one species is not reflected at the generic level. Thus, a major extinction event that resulted in the loss of 60% of the species extant in a region might be reflected in the loss of only 30% of the genera or 10% of the families represented.

B. Background and Mass Extinctions

Since the origin of life approximately 3.5 billion years ago, extinction has been a major process influencing the course of its evolution. Current estimates from the fossil record indicate that more than 90% of the species that once existed are now extinct; some suggest that the number is more than 99%. Many workers, particularly J. J. Sepkoski Jr., D. M. Raup, and their associates (Raup, 1991), have contributed greatly to our understanding of the contribution of extinction to patterns of change in diversity. The majority of these studies have been based on extensive databases of records of the times of origin and extinction of families and genera of marine invertebrates and vertebrates compiled by Sepkoski. Temporally, these records of extinctions (including both lineage extinctions and pseudoextinctions) are ordered on a scale of geological stages. These stages vary in length but have an average duration of about 6 million years, an interval about three times as long as the Quaternary. This difference in degree of resolution must be emphasized in attempts to apply the results of studies of pre-Quaternary extinctions to analyses of Quaternary, particularly modern, patterns of extinction.

Sepkoski's initial studies indicated that rates of extinction fell into two categories, which were termed background extinctions and mass extinctions. Background extinction rates appeared to characterize the majority of the history of life. Several times during the past 600 million years rates of extinction increased greatly over short periods of geological time; these were recognized as mass extinction events. Subsequent research led Raup to develop what he termed the "kill curve," which highlighted the fact that rates of extinction varied over a considerable spectrum with background extinctions and mass extinctions merely representing the extremes. His analysis, however, did not negate the observation that during the past 600 million years the pattern of evolution of marine organisms was greatly modified by mass extinctions.

The most severe of the mass extinction events frequently are identified as the "big five" mass extinctions. Although the available database for evolution of terrestrial organisms is not as detailed as that compiled by Sepkoski for the marine record, enough is known to show that some of these mass extinctions modifying the marine biota had significant impacts on the evolution of terrestrial life. In particular, the mass extinctions used to mark the Permian–Triassic and Cretaceous–Tertiary boundaries significantly influenced the patterns of evolution of "nonmammalian synapsids" and early mammals.

C. Temporal Scale of Mass Extinctions

Mass extinctions were intentionally defined subjectively as short periods of Earth history during which rates of extinction reached exceptionally high levels. In geological terms, how short is short? In studies of recent extinctions in which events are followed on an ecological timescale, short periods of time are measured in terms of years, decades, centuries, or possibly a millennium or two. Current radiometric methods for determining ages and durations of pre-Quaternary events lack this level of resolution. For example, using the $^{40}Ar/^{39}Ar$ method of age determination, which is the most precise method of radiometric age determination available for pre-Quaternary deposits, the age of the Cretaceous–Tertiary boundary is placed at 65.16 ± 0.04 Ma. As data on extinctions of Cretaceous lineages of marine and nonmarine organisms are collected and correlated, all extinction events that occurred within an interval of 80,000 years would necessarily be treated as having occurred simultaneously. The farther one goes back in Earth history, the longer the error bars become; for example, some paleontologists currently argue that the extinctions of lineages that comprise the Permian–Triassic mass extinction might have occurred over an interval of 1 or 2 million years.

D. Distal and Proximal Causal Factors

In studies of extinction it has proven useful to distinguish between distal and proximal causal factors. Proximal causal factors are those that impinge directly on individuals and are the immediate causes of their death. Distal causal factors are the regional or global events that result in changes in local (proximal) physical and biological environments affecting individual organisms. As a current example, the various sources of atmospheric pollution thought to contribute to global warming would be ranked as distal factors. The impacts of changes in annual temperature regime on individuals would be proximal factors.

In the much longer period of mammalian evolution prior to the Quaternary, several distal causal factors emerge as major contributory factors to mass extinctions and other significant changes in the course of evolution of the group. Recognition of some of these can be traced to the development of our understanding of plate tectonics, the mechanisms involved in movements and changes in configurations of continents and oceanic basins. Immediate consequences of these events have been the modification of patterns of circulation of oceanic currents and continental topography, which contributed to climatic changes. Hallam and Wignall (1997) and other workers have noted that many of the major mass extinctions were associated with times of withdrawal of shallow continental seas (marine regressions) and reduction of coastal habitats.

Another product of changes in configurations of continents during periods of marine regression is the establishment of terrestrial connections between long-isolated continents. Two recent examples of this process are the formation of dry land connections between Siberia and Alaska and between North and South America via the Panamanian Isthmus. Both facilitated mammalian interchange and extinction of some lineages, probably through some combination of competition, predation, and introduction of new diseases.

Volcanic activity has played and continues to play a significant role in modifying the earth's environments. Major eruptions produce large quantities of particulate matter, aerosols, and acids. These clouds can reach magnitudes sufficient to limit or block the sun's radiation from reaching the surface of the earth, causing a decrease in temperature. Later, they would trap heat radiated from the surface to produce a greenhouse warming of the climate. Increased amounts of acid rain also are predictable consequences of volcanic activity.

Although the hypothesis had precursors in earlier decades, the assertion by Alvarez *et al.* (1980) that the impact of an asteroid was the causal factor of the Cretaceous–Tertiary mass extinction focused attention on the potential role of extraterrestrial bombardment in controlling the course of evolution of the earth's biota. Their initial hypothesis argued that the impact of an asteroid was the distal cause of the extinction of nonavian dinosaurs and numerous other kinds of terrestrial and marine organisms. The hypothesized proximal cause was the formation of dust clouds that first produced global refrigeration by blocking solar energy and then a greenhouse effect by trapping heat in the atmosphere. Acid rain and extensive wildfires also were suggested as contributory factors. Qualitatively, these are the same kinds of perturbations of the environment that could be generated by intense volcanic activity. As discussed later, the pattern of survival and extinction of mammals across the Cretaceous–Tertiary boundary has played a significant part in testing the impact hypothesis and other hypotheses concerning the causal factors of this mass extinction.

It has been argued that a mass extinction is an exceptional event and, therefore, requires an exceptional causal factor. Two aspects of this assertion require comment. Analyses of the geological and fossil records indicate that many of the potential distal causal factors, for example, regressions of the seas, outbursts of volcanic activity, and impacts of large extraterrestrial bodies, were contemporaneous with mass extinctions and lesser changes in the earth's biota. None of them has been demonstrated to be contemporaneous with all the major mass extinctions. Examples of each of these potential causal factors have been recorded at times when mass extinctions in the marine and/or terrestrial biota have not occurred. Now, some scientists suggest that the causes of mass extinction are to be found in combinations of these and possibly other factors that compounded to produce major, deleterious changes in the environment.

Second, the level of ecological interdependence of members of the terrestrial biota has, no doubt, varied through time. Terrestrial ecosystems characterized by relatively low levels of interdependence of their members might well be able to withstand the loss of one or a few species. In contrast, ecosystems with highly interdependent members might not be able to withstand the loss of one or a few species, resulting in the collapse of the whole system and consequent extinction of many other species. Thus, a period of global cooling, for example, could have a catastrophic effect on highly interdependent ecosystems but cause little or no change in those that were characterized by a lower level of interdependence.

Finally, the focus of this article is extinctions of lineages and their effects on patterns of evolution and diversity. Extinction is but one-half of the equation for calculating changes in biodiversity. Variation in the rate of origination of new species can have a marked effect on biodiversity. For example, analysis of the fossil record of nonavian dinosaurs during the Jurassic and Cretaceous shows that they evolved rapidly; a genus of nonavian dinosaur rarely lasted more than a geological stage (on average 6 million years in duration) or two. Throughout most of their history, rates of extinction and origination of nonavian dinosaurs were high. In the last two stages prior to their extinction at the end of the Cretaceous, a high rate of extinction of genera appears to have been maintained, but the rate of origination of new genera was depressed.

III. ORIGIN OF SYNAPSIDS AND THE PERMIAN–TRIASSIC MASS EXTINCTION

During the Carboniferous Period (Table I), approximately 354–290 Ma, the reptilian and synapsid clades differentiated. From this beginning, the first phase of synapsid evolution, characterized by their prominent position in terrestrial faunas, would last until the end of the Triassic. It was interrupted, and the course of synapsid evolution altered, by the massive extinctions that mark the Permian–Triassic boundary.

In the Early Permian, basal "nonmammalian synapsids," the "pelycosaurs," were the most diverse terrestrial vertebrates. Notable within the evolutionary radiation of "pelycosaurs" were two lineages that evolved long extensions of the neural spines of their vertebrae, which supported high dorsal fins. These fins, which greatly increased the surface area of the body but not its volume, have been interpreted as early experiments in thermoregulation. Exposed to the sun, the fin would have sped up heating of the body; in the shade it would have served as an efficient radiator of heat.

The "pelycosaur"-dominated terrestrial faunas of the Early Permian were composed of many more carnivores than herbivores; at first glance, this appears to be an ecologically anomalous situation. It is now thought that the dominance of carnivores is not a product of a bias in preservation or collecting but reflects an initial stage in the evolution of the terrestrial ecosystem. Many of the species of these early "nonmammalian synapsids" may well have been amphibious, returning to streams and lakes to feed on fish and freshwater invertebrates. In about the middle of the Permian, the composition of terrestrial faunas changed with the appearance of a diversity of "therapsids" (Fig. 1). Herbivourous forms, including the dicynodonts, were much more abundant, which suggests that fully terrestrial ecosystems had evolved.

The mass extinction used to mark the end of the Permian was the most severe yet recorded. It has been estimated that more than 90% of the species and 50% of the families of marine invertebrates were lost during this event. Terrestrial ecosystems also were greatly modified by extinctions. Of the approximately 50 families of terrestrial vertebrates present in the past 5 million years of the Permian, approximately 75% died out. This dramatic loss included large and small herbivores and carnivores. Notable survivors among the "nonmammalian synapsids" were a few lineages of carnivorous "therapsids" and the herbivorous dicynodonts.

Debate continues over the causal factors of the mass extinction marking the end of the Permian (Hallam and Wignall, 1997). A variety of lines of evidence suggest that this was a time of global warming, probably related to a marine regression followed by development of anoxic oceans. Extensive volcanic activity, the eruption of the massive flood basalts of western Siberia, was probably another distal causal factor. Currently, students of the Permian–Triassic mass extinction appear to favor some combination of changes in the oceans and volcanic activity as its distal causal factors. No evidence of the impact of an asteroid or other extraterrestrial body has been discovered.

IV. THE SECOND "THERAPSID" RADIATION AND LATE TRIASSIC FAUNAL TURNOVER

As a result of the decimation caused by the Permian–Triassic mass extinction, the taxonomic diversity of earliest Triassic terrestrial faunas was greatly reduced. These faunas were dominated by a dicynodont, *Lystrosaurus*, but included representatives of a few other lineages of "nonmammalian synapsids," which were the basal stocks of another evolutionary radiation. Later, in the Early Triassic, numerous large and small herbivorous forms, primarily dicynodonts, became prominent elements of terrestrial faunas. In parallel, the surviving carnivorous "therapsids" radiated, producing a variety of new lineages of large and small carnivores as well as some herbivorous species.

Analysis of the evolutionary radiation of the carnivorous "therapsids" during the Triassic reveals trends in modification of skull structure and morphology of the postcranial skeleton that increasingly resemble the characteristics of modern mammals. For example, the structure of the lower jaw was modified through expansion of the dentary bone and reduction or losses of other bones characteristic of early synapsid or reptilian jaw structure. These were preliminary steps toward acquisition of an articulation of the dentary with the squamosal bone of the skull, a key osteological character for distinguishing modern mammals from reptiles. Limb posture began to be modified from a primitive, sprawling stance to a more typically mammalian stance with the elbow and knee tucked close to the body and the limbs brought into an upright posture.

A few million years before the end of the Triassic, some synapsids had evolved a functional articulation between the dentary bone of the lower jaw and the squamosal. Many textbooks and popular articles present the acquisition of this structure as the hallmark of the Mammalia and identify the end of the Triassic (approximately 206 Ma) as the time of origin of the group. In current phylogenetic classifications, however, the clade Mammalia is restricted to a crown group including the common ancestor of living monotremes (the echidna and platypus), marsupials, and placentals and all its descendants. Phylogenetically, some of the very advanced "nonmammalian synapsids" with a dentary–squamosal jaw articulation and other mammal-like specializations lie outside the crown group; they are sister lineages. In many recent research papers, these very advanced "nonmammalian synapsids" are referred to as "mammaliaforms."

During the Late Triassic, the composition of terrestrial faunas underwent a major change that played a significant role in reshaping the course of synapsid evolution. All the lineages of Triassic synapsids, except for a few lineages of "mammaliaformes" and one group of herbivorous "nonmammalian synapsids," the tritylodonts, became extinct. All the surviving lineages were represented by species of relatively small body size. Two groups of reptiles, the rhynchosaurs, and several lineages of dinosaurs became the dominant terrestrial vertebrates.

Recent studies indicate that the Triassic–Jurassic mass extinction's primary impact was on the marine biota. Some evidence suggests that terrestrial floras were modified at this time. The admittedly limited fossil record of Late Triassic and Early Jurassic "nonmammalian synapsids" and "mammaliaforms" suggests that these groups were not greatly affected. Clearly, the radiation of "mammaliaforms," which began in the Late Triassic, continued in the Jurassic.

V. MAMMALIAN EVOLUTION DURING THE AGE OF DINOSAURS

The extensive changes in terrestrial faunas during the Late Triassic brought the first phase of synapsid evolution to a close. During the Jurassic and Cretaceous, popularly dubbed the Age of Dinosaurs, "mammaliaforms" and early mammals were not dominant members of terrestrial faunas. For the most part, they were very small animals, in the size range of the smallest modern mammals. A few evolved larger body size, rivaling modern opossums or raccoons, but these species were exceptions to the rule.

During the Jurassic, "mammaliaforms" differentiated. Functional interpretations of their dentitions suggest that most were probably carnivorous, feasting on small prey including small terrestrial vertebrates and invertebrates. One group, the multituberculates, evolved a remarkably rodent-like style of dental specialization and is interpreted as having been omnivorous or herbivorous in dietary preferences. This interpretation is strengthened by the persistence of the group into the Paleocene. Then, as rodents began their evolutionary radiation in the late Paleocene, the taxonomic diversity of multituberculates dwindled. By the end of the Oligocene, when almost all families of rodents had differentiated, multituberculates were extinct.

Tracing the course of mammalian evolution during this second phase of synapsid evolution is complicated by distinct biogeographic differentiation. Unlike the Permian and Triassic, when most of the terrestrial regions of the world were parts of one supercontinent, Pangaea, through the Jurassic and Cretaceous terrestrial areas and their faunas began to be fragmented. In part, this was the product of plate tectonic processes that broke up the Pangaean supercontinent and shifted the positions of the resulting continental blocks. Additionally, the Age of Dinosaurs was a time characterized by extensive marine transgressions when shallow seas covered many areas and added to fragmentation of terrestrial areas. By the beginning of the Cretaceous, when the currently sparse fossil record of these faunas gives us our first real picture of biogeographic differentiation, there is evidence of distinct northern (North American–Eurasian) and southern (Australian–Antarctic–South American) terrestrial faunas. Unfortunately, very little is known of the history of "mammaliaform" and mammalian evolution on the African continent during this interval.

The Cretaceous was a time of major change in the terrestrial biota. Angiosperms made their appearance and began to diversify. First records of most modern families of insects document their diversification. Particularly in the Northern Hemisphere, dinosaurian faunas underwent a major reorganization. Global climates of the Jurassic and Cretaceous were, in general, distinctly warmer and more equable than modern climates. The latitudinal climatic gradient, reflecting the degree of difference between equatorial and polar conditions, was much lower. Throughout most of the interval there is no evidence of glaciation. Finally, to round out this description of the setting of mammalian evolution during the Age of Dinosaurs, it must be noted that evolution of the terrestrial biota was not interrupted by a mass extinction event.

By the middle of the Cretaceous, the mammalian fauna had undergone a major turnover in composition. In the Northern Hemisphere, with the exception of the hardy multituberculates, most of the Jurassic lineages of "mammaliaforms" had dwindled in taxonomic diversity or become extinct. Concurrently, two of the three major lineages of the crown group Mammalia, marsupials and placentals, appeared and began to radiate.

Mammalian evolution on a southern continent, composed of the modern Australian, Antarctic, and South America continents, followed a different course. The terrestrial fauna of this area includes the first records of the third major lineage of the crown group Mammalia, the monotremes. Additionally, at least two groups of "mammaliforms" diversified there.

Currently, the first records of the crown group Mammalia are of Early Cretaceous age. The biogeographic dichotomy, the first appearances of marsupials and placentals in the Northern Hemisphere and the first appearance of monotremes in the southern continent, strongly suggests that the first records of the crown group Mammalia will be found in the Jurassic.

VI. CRETACEOUS–TERTIARY MASS EXTINCTION AND ITS CONSEQUENCES

The effects of the Permian–Triassic mass extinction only overshadow the extent of devastation of marine and terrestrial biotas during the Cretaceous–Tertiary mass extinction. On land, the Cretaceous–Tertiary mass extinction is marked by the demise of nonavian dinosaurs and some other vertebrate lineages. Similarly, many dominant groups of marine invertebrates and aquatic reptiles died out at this time.

The primary source of information documenting the effects of the Cretaceous–Tertiary mass extinction on a terrestrial biota is a series of studies of faunal and floral evolution in an area of the Western Interior of North America that extends from Colorado northward into Alberta, Canada. For decades, a wide range of geological and paleontological research projects have been under way in this area, and the resulting database is extensive. This is both a blessing and a curse. There has been a tendency to regard the results of studies on biotic change in the North American Western Interior as typical of global patterns. Recent discoveries in other areas show that this was not the case. On the other hand, although geographically limited, the extensively documented patterns of survival and extinction in the North American Western Interior can be used to test hypotheses concerning the kinds and severity of environmental changes that caused the Cretaceous–Tertiary mass extinction.

Archibald (1996) and others have analyzed the patterns of survival and extinction of terrestrial vertebrates based on extensive collections from northeastern Montana. In the Cretaceous–Tertiary mass extinction all the nonavian dinosaurs, approximately 20 genera, became extinct. A small group of freshwater sharks and their relatives shared the fate of the nonavian dinosaurs. These losses, plus extensive extinctions among the Cretaceous lineages of lizards and marsupials, account for almost 75% of the extinctions of terrestrial vertebrate species in the North American Western Interior. In contrast, at the species level, no extinctions of lineages of frogs, salamanders, and placental mammals have been documented. Only a few species of turtles, crocodilians, and crocodilian-like reptiles died out at this time.

Since Archibald published his analysis, several paleontologists have begun to analyze the very fragmentary fossil record of the evolution of birds in the Late Cretaceous. Their research shows that the evolutionary diversification of the crown group of birds had begun in the Cretaceous. Although many lineages of more primitive birds died out before or at the Cretaceous–Tertiary boundary, extinction appears to have taken little toll among the lineages of the crown group.

Among the mammalian lineages present in the North American Western Interior, approximately 50% of the species of "mammaliaform" multituberculates, which are thought to be the ecological equivalents of rodents, became extinct during the Cretaceous–Tertiary mass extinction. All lineages of placental (also dubbed eutherian) mammals survived this event. In stark contrast, all but one of the Cretaceous lineages of marsupials appear to have become extinct at this time.

Recent discoveries in Australia and, particularly, South America are beginning to outline the course of mammalian evolution through the Cretaceous–Tertiary mass extinction in the Southern Hemisphere. At this time, Australia, Antarctica, and South America were closely approximated and global climates were much warmer. Direct evidence documents interchange of terrestrial vertebrates between Australia and South America. In South America, deposits laid down several million years prior to the end of the Cretaceous contain records of a mammalian fauna dominated by members of several lineages of "mammaliaforms" accompanied by a lineage of monotremes. A gap in the fossil record separates records of this fauna from deposits documenting the earliest Tertiary fauna known from the area. The change in mammalian faunal composition is striking. Most of the "mammaliaform" lineages had gone extinct. The lineage of monotremes survived into the Tertiary but soon became extinct. The dominant group of mammals in the earliest Tertiary of South America is a small but diverse group of marsupials that were accompanied by a few lineages of placental mammals. These marsupials and placentals appear to be derived from stocks that dispersed southward from North America about the time of the Cretaceous–Tertiary boundary.

Other continents have yet to yield extensive records of mammalian evolution across the Cretaceous–Tertiary boundary. The available data strongly suggest that the mass extinction at this time involved the termination of mammalian lineages in these areas. Globally, mammalian diversity appears to have been greatly decreased.

The distal causal factors of the Cretaceous–Tertiary mass extinction are the subject of continuing debate. The pattern of survival and extinction of species of terrestrial vertebrates at the Cretaceous–Tertiary boundary in the North American Western Interior argues strongly against, if not falsifies, hypotheses calling for massively catastrophic changes in the environment as distal causal factors of this mass extinction. Discovery of a large crater in Yucatan adds evidence supporting the hypothesis that an asteroid, or other extraterrestrial body, impacted the earth at the close of the Cretaceous. Additionally, massive volcanic deposits in peninsular India document a period of extensive eruptions that began before and continued across the Cretaceous–Tertiary boundary. Finally, the end of the Cretaceous is a time of a major marine regression. Certainly, all three of these events contributed to changes in global environments that triggered the mass extinction; their relative importance remains an open question (Archibald, 1996; Hallam and Wignall, 1997).

VII. RECOVERY OF MAMMALIAN FAUNAS DURING THE EARLY PALEOCENE

The bias favoring the fossil record from the Northern Hemisphere continues to limit analyses of the course of mammalian evolution immediately after the Cretaceous–Tertiary extinctions, which marked the end of the second phase of synapsid evolution. Earliest Paleocene faunas are well-known only in the North American Western Interior. In northeastern Montana and immediately adjacent areas, a refined fossil record provides the basis for study of the recovery of the mammalian fauna during the first million years of the Paleocene (approximately 65–64 Ma). Radiometric age determinations make it possible to subdivide this period of faunal recovery into a first interval of approximately 400,000 years and a second of approximately 600,000 years.

In the first 400,000 years of the Paleocene, diversity of the mammalian fauna of the North American Western Interior was depressed in comparison to that of the latest Cretaceous. The most numerous mammals were placentals, primarily represented by the paraphyletic "condylarths." This group was allied to lineages of modern ungulates, hoofed mammals, that would make their appearance later in the Paleocene or in the Eocene. The second most diverse and abundant group was the "mammaliaform" multituberculates. Only one lineage of marsupials remained. All these earliest Paleocene mammals were characterized by small body size. As far as can be determined from their fragmentary remains, none was larger than the largest latest Cretaceous mammals known from the area, which were about the size of modern opossums or raccoons.

Less than half of the mammalian species in this earliest Paleocene fauna represent lineages present in the preceding latest Cretaceous fauna of the Western Interior. The majority is composed of immigrants that appear to have differentiated in other areas and dispersed into the Western Interior after the Cretaceous–Tertiary mass extinction. This is one line of evidence that suggests the diversification of major placental lineages had begun in other, as yet unsampled, areas prior to the end of the Cretaceous.

Support for this view also comes from recent comparative molecular studies of modern placental mammals. Through application of the molecular clock technique, these have produced hypotheses suggesting very ancient times of differentiation of major placental lineages. The molecular clock technique is based on the assumption that the rate of accumulation of molecular differences has been relatively regular. If this is correct, the time of differentiation of two lineages can be estimated on the basis of their current degree of molecular difference. In large part, these molecular studies suggest that the differentiation of many major lineages of placentals took place in the middle or early part of the Cretaceous, if not earlier. This conclusion directly contradicts the current fossil record, which suggests the early differentiation of these lineages occurred in the Late Cretaceous or was a product of the rapid evolutionary radiation of placentals in the Paleocene. Probably the true history of placental evolution is to be found somewhere between these extremes (Novacek, 1999).

Regarding charting the course of mammalian recovery after the Cretaceous–Tertiary boundary, during the next 600,000 years of the Early Paleocene there was a major increase in mammalian, particularly placental, diversity in the Western Interior. In part, this increase was the result of diversification of lineages present in the area at the beginning of the Paleocene, but immigration of additional groups of placentals and multituberculates from other areas played a significant role. During this interval, body size increased in several lineages. For the first time since the Late Triassic, placentals and multituberculates achieved sizes that approximated those of modern middle- and large-sized mammals. These trends of an increase in taxonomic diversity and the appearance of species of large body size continued throughout the later part of the Paleocene.

VIII. MAMMALIAN EVOLUTION DURING THE TERTIARY

From the late Paleocene onward, mammalian faunas of the Northern Hemisphere are increasingly well documented by assemblages of fossils from both Eurasia and North America. After the initial recovery of the mammalian fauna of the Northern Hemisphere from the effects of the Cretaceous–Tertiary mass extinction, diversity continued to increase, and many lineages were characterized by evolution of increasingly larger individual body size. Before the end of the Paleocene some species had attained the bulk of modern rhinos or hippos.

The late Paleocene and Eocene was a period of marked faunal turnover in the Northern Hemisphere, with the appearance of an increasing number of lineages of placentals that are represented in modern faunas; for example, this was the time of appearance of the equids and other kinds of modern ungulates. Additional contributions to mammalian diversity came with the first records of flying mammals (bats) and the appearance of one group of marine mammals (whales). Evolution of the marine seals and sea lions occurred later, in the Oligocene. Also, there were instances of decreases in diversity or extinction of major lineages. For example, during the late Paleocene and Eocene the diversity of the "mammaliaform" multituberculates decreased with the appearance and diversification of rodents. The correlation suggests that competition with rodents, and possibly other small, placental herbivores, played a significant role in the demise of the multituberculates. In general, however, the diversity of terrestrial mammals increased.

Near the Eocene–Oligocene boundary, approximately 33.7 Ma, many mammalian lineages became extinct. In comparison to other extinctions, this was not a mass extinction event, but it significantly modified the composition of mammalian faunas of the Northern Hemisphere. The extinctions occurred at about the time of a marked cooling in global climates and were most frequent in groups of placentals thought to have been particularly adapted to tropical forests, which were widespread in the Eocene. In addition to having been cooler, Oligocene climates were characterized by a general increase in aridity and decrease in equability, which was reflected in the evolution of greater areas of savanna and steppe. Analysis of patterns of evolution of mammalian body size during the Eocene and Oligocene suggest a polarization with increases in diversity of species of large and small body size but a decrease in diversity of species of intermediate size.

Mammalian diversity in the Northern Hemisphere, measured at the family levels, reached a maximum in the early Miocene, particularly reflecting a radiation of several lineages of ungulates. Through the remainder of the Miocene and Pliocene, the diversity of large ungulates decreased as the climate became cooler and dryer. Although differing in detail, broadly similar patterns in the incidence of extinctions and faunal turnover characterize the evolutionary history of mammalian faunas of the continents of the Southern Hemisphere.

Until recently, climatic changes were thought to be the major, if not the only, driving force in mammalian evolution during the Tertiary. During the Paleocene global climates became distinctly warmer and reached maximum levels in the Eocene, when tropical forests extended into high latitudes. During the transition from the Eocene into the Oligocene, global climates appear to have changed rapidly, on a geological timescale. This was a time of extinction of many lineages of terrestrial and marine mammals. Although at a more gradual rate, cooling of the environment continued up to the beginning of the Quaternary and inception of major continental glaciation. The circumstantial evidence of a correlation of periods of mammalian diversification or extinction with patterns of climatic change has been challenged by recent studies (Alroy, 1998; Prothero, 1999). A variety of high-resolution paleontological and geological methods were used to describe and test these correlations. The results indicated that correlations of times of climatic change and periods of increased diversification or extinction were not precise. Other environmental or biological factors apparently contributed to controlling the course of Tertiary mammalian evolution.

IX. CONCLUSIONS

The approximately 300-million-year history of synapsid evolution can be subdivided into three major phases. During the first phase, "nonmammalian synapsids" were among the dominant terrestrial vertebrates. The second phase, which began in the Late Triassic and lasted until the end of the Cretaceous, was characterized by the dominance of the dinosaurs in terrestrial faunas. "Mammaliaform" synapsids and early mammals were a lesser component of these faunas. The third phase began after the Cretaceous–Tertiary mass extinction and was characterized by a major diversification of mammals, returning them to a dominant position.

In addition to the ongoing effects of background extinction, mass extinctions played a significant role in directing the course of synapsid evolution. The Permian–Triassic and, particularly, the Cretaceous–Tertiary mass extinction events greatly reduced synapsid diversity and reset the course of its evolution. Other major turning points in the evolution of the group, which involved extensive extinctions of lineages, were not coincident with mass extinctions—for example, the periods of faunal turnover late in the Triassic and during the transition from the Eocene into the Oligocene.

Attempts to compare patterns and causal factors of pre-Quaternary and Quaternary mammalian extinctions must consider the significant differences in resolution of the timescales at which these events can be studied. In some instances, the fossil record of Quaternary, particularly later Quaternary, extinctions permits analysis of biological change on ecological timescales of hundreds or thousands of years. Resolution of the pre-Quaternary record is much less. The severity of these extinction events may be artificially compounded by our inability to distinguish events that occurred at different times over intervals of hundreds of thousands or a few million years.

Finally, a survey of current research on the distal causal factors of mass extinctions and periods of rapid faunal turnover in synapsid evolution indicates that there probably was not a single grim reaper. Varying combinations of the effects of modification of the configuration of oceans and their patterns of circulation, increases in the intensity of volcanic activity, and the sporadic impacts of extraterrestrial bodies all contributed to sometimes deleterious changes in the physical environment. Similarly, biotic factors played a part. The degree of interdependence of the members of an ecosystem, for example, contributes to the system's resistance to environmental change. Competition and predation also have contributed to triggering extinctions. At different times in the pre-Quaternary history of mammalian evolution, all these and probably other factors caused extinctions of lineages.

See Also the Following Articles

DINOSAURS, EXTINCTION THEORIES FOR • EXTINCTION, CAUSES OF • MAMMALS, BIODIVERSITY OF • MAMMALS, LATE QUATENARY, EXTINCTION OF • MASS EXTINCTIONS, CONCEPT OF • MASS EXTINCTIONS, NOTABLE EXAMPLES OF

Bibliography

Alroy, J. (1998). Long-term equilibrium in North American mammalian diversity. In *Biodiversity Dynamics: Turnover of Populations, Taxa, and Communities* (M. L. McKinney, Ed.), pp. 232–287. Columbia University Press, New York.

Alvarez, L. W., Alvarez, W., Asaro, F., and Michel, H. (1980). Extraterrestrial cause for the Cretaceous–Tertiary extinction. *Science* 208, 1095–1108.

Archibald, J. D. (1996). *Dinosaur Extinction and the End of an Era: What the Fossils Say*. Columbia Univ. Press, New York.

Benton, M. J. (1990). *Vertebrate Paleontology*. Chapman & Hall, London.

Hallam, A., Wignall, P. B. (1997). *Mass Extinctions and Their Aftermath*. Oxford Univ. Press, Oxford.

Kemp, T. S. (1982). *Mammal-like Reptiles and the Origin of Mammals*. Academic Press, London.

Novacek, M. J. (1999). 100 million years of land vertebrate evolution: The Cretaceous–Early Tertiary transition. *Ann. Missouri Bot. Garden* 86, 230–258.

Prothero, D. R. (1999). Does climatic change drive mammalian evolution? *GSA Today* 9 (9), 1–7.

Raup, D. M. (1991). *Extinction: Bad Genes or Bad Luck?* Norton, New York.

MANGROVE ECOSYSTEMS

Peter J. Hogarth
University of York

GLOSSARY

aerenchyma A spongy plant tissue composed largely of air spaces enabling gas exchange to take place by diffusion in underground mangrove roots.

aerial roots In mangrove species such as *Rhizophora*, roots branch out from the stem some distance above the soil surface. Lenticels (pores) in the aerial portion of these roots enable gas exchange to take place, through aerenchyma tissue, with the respiring underground portions of the root.

mangal A term sometimes used to specify the mangrove habitat as a whole as opposed to "mangrove" applying specifically to the trees themselves. For the most part, however, mangrove is considered to apply to both trees and habitat.

pneumatophores In some species of mangrove, such as *Avicennia* and *Sonneratia*, underground roots spread laterally from the main stem. Pneumatophores grow vertically from these, typically standing 10–20 cm above the soil surface, enabling gas exchange to take place with the underground roots.

pseudofecal pellet Fiddler crabs and their relatives collect soil with their mouthparts, separate organic particles from mineral components by a complex flotation process, ingest the former, and discard the latter in the form of compact pellets. These are known as pseudofecal because, although extraction has taken place, the waste material has not passed through the gut.

MANGROVES ARE a group of trees and shrubs, mostly evergreen, which have convergently evolved physiological and morphological adaptations to shallow intertidal environments. These are mostly composed of soft sediment, in which other vascular plants are rare. Mangroves are almost exclusively tropical in distribution and often dominate large areas of coastline or estuary.

I. MANGROVE TREES

Currently, the total mangrove area in the world is estimated at 170,000 km^2. They are the principal source of primary productivity in such areas. By their presence, they also provide shelter for other organisms. Mangroves are therefore the energy base, and physical substrate, of an often complex and diverse ecosystem. Mangrove faunas, to a unique extent, comprise organisms of both marine and terrestrial origin.

Encyclopedia of Biodiversity, Volume 3

A. The Mangrove Habitat

The mangrove environment is a demanding one. Typically, mangroves are regularly inundated by tides and are therefore usually in a permanently waterlogged state. The tidal water is saline, so mangrove trees have the problem of coping with salt and acquiring sufficient water against an osmotic gradient. In hot climates, evaporation may make the salinity even greater than that of seawater. In the Indus Delta (Pakistan), for example, the prevailing salinity may be as much as twice that of seawater. Among the vascular plants, only mangroves flourish in such an inhospitable environment (Fig. 1).

Mangroves are defined physiologically as trees that can survive in the mangrove habitat, or mangal. The term is not a taxonomic one, nor does it indicate phylogenetic divergence from a common mangrove ancestor. The approximately 50 species generally recognized as mangroves belong to 20 genera in 16 families, although 2 families, Avicenniaceae and Rhizophoraceae, dominate in number of species (as they do also in abundance) (Table I). In most cases these genera and families also contain nonmangrove members. Mangrove species have evolved their specialist features as the result of convergent evolution, and mangrove attributes have probably evolved independently at least 15 times.

In addition to true mangrove species, there is also a loosely defined category of mangrove associates. These are species often occurring in mangrove habitats but which also occur elsewhere. Some are found only at the landward margins of the mangal, whereas others, such as creepers and lianes, have their roots above the intertidal zone but invade the mangal by using the mangrove trees purely for support. Other plants associated with mangrove trees are epiphytes, which include ferns and the "ant-house" plants (see Section II.A), and parasitic mangrove mistletoes.

B. Adaptations to the Mangrove Environment

1. Salinity

Three principal mechanisms enable mangrove trees to survive saline environments. Some species exclude salt at the root surface while continuing to take in water. In *Aegiceras* and *Avicennia*, up to 97% of the salt is excluded, apparently by a physical rather than a metabolic mechanism. This has the effect of locally increasing the salinity of the soil around the roots, with implications for other organisms: mangrove trees modify their environment as well as respond to it. In other instances, trees take in salt but sequester it within cells in such a way that sensitive metabolic processes are protected from contact with excessive salt concentrations. Finally, several mangrove species secrete excess salt, at considerable metabolic cost, from specialized salt glands on their leaves. Many mangrove species use a combination of these mechanisms, as shown in Table II.

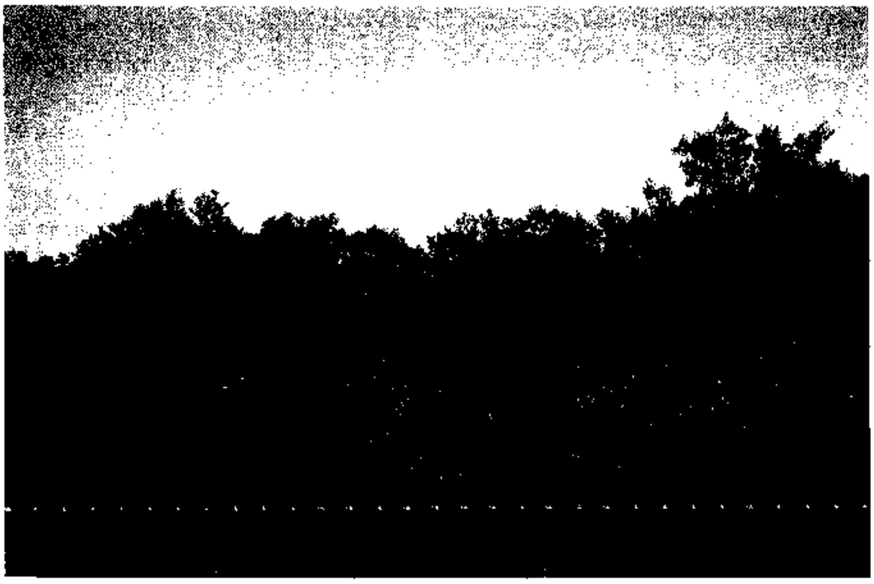


FIGURE 1 Mangroves (*Avicennia* and *Rhizophora*) fringing a tidal creek in the Indus Delta, Pakistan.

TABLE I

Distribution of Mangrove Species by Family and Genus[a]

Family	Genus	Number of mangrove species
Avicenniaceae	*Avicennia*	8
Combretaceae	*Laguncularia*	1
	Lumnitzera	2
Palmae	*Nypa*	1
Rhizophoraceae	*Bruguiera*	6
	Ceriops	2
	Kandelia	1
	Rhizophora	5
Sonneratiaceae	*Sonneratia*	5
Bombacaceae	*Camptostemon*	2
Euphorbiaceae	*Excoecaria*	1
Lythraceae	*Pemphis*	1
Meliaceae	*Xylocarpus*	2
Myrsinaceae	*Aegiceras*	2
Myrtaceae	*Osbornia*	1
Pellicieraceae	*Pelliciera*	1
Plumbaginaceae	*Aegialitis*	2
Pteridaceae	*Acrostichum*	3
Rubiaceae	*Scyphiphora*	1
Sterculaceae	*Heritiera*	3
Total 16	20	54

[a] This follows the classification of Tomlinson (1986); there are althernative views on the status of certain species as true mangroves or mangrove associate species.

TABLE II

Methods of Salt Tolerance Employed by Mangrove Species

Species	Exclude	Secrete	Accumulate
Acanthus		+	
Aegialitis	+	+	
Aegiceras	+	+	
Avicennia	+	+	+
Bruguiera	+		
Ceriops	+		
Excoecaria	+		
Laguncularia		+	
Osbornia	+		+
Rhizophora	+		+
Sonneratia	+	+	+
Xylocarpus			+

2. Waterlogging

The major problem of waterlogged soils is lack of oxygen. Underground roots, like all tissues, require oxygen for respiration. In a normal soil, gas exchange takes place readily through air-filled spaces between soil particles. In water, the rate of diffusion of oxygen is very low, and in consequence waterlogged soils are generally virtually lacking in free oxygen. One of the most widespread mangrove trees, *Rhizophora*, adapts to such anoxic soils by keeping much of the root mass above the mud surface, surrounded by air. The stretches of these aerial roots (Fig. 2) close to the soil carry numerous gas-exchange pores, or lenticels, whereas the underground portions are honeycombed with air-filled spaces.

This air-filled tissue, or aerenchyma, is also a feature of *Avicennia* and *Sonneratia*, whose roots are horizontal and close to the surface. These species respire by growing numerous pencil-like pneumatophores which protrude above the mud surface and allow gas exchange with the underground tissues (Fig. 3). Pneumatophore growth is facultative: The less waterlogged the soil, the lower the pneumatophore density. In the extreme and atypical case of *Avicennia* growing in sand between the Egyptian Sinai desert and the sea, the soil is so well oxygenated that no pneumatophores develop.

The aerial roots of *Rhizophora* and the intertwined underground horizontal roots of *Avicennia* physically support the trees in what is often a relatively unstable and shifting soil. Aerial roots and pneumatophores provide attachment sites for epibionts and facilitate the accretion of sediment by impeding water movement.

3. Reproduction

Many mangrove species show some form of vivipary. *Rhizophora* is an example. The ovum is fertilized while still on the parent tree and grows by a combination of photosynthesis and acquisition of nutrients from the parent until it may reach a length of 50 cm (Fig. 4). This structure—neither a seed nor a fruit, and hence usually termed a propagule—then falls to the ground. The propagules of some species root almost immediately, but others appear to have an obligatory floating period before they sink and establish themselves. The majority of floating propagules probably settle close to the parent, but long-distance dispersal is also possible. Floating mangrove propagules may remain viable for a month or longer: Depending on current speed and

FIGURE 2 Aerial roots of *Rhizophora* in a Malaysian mangrove forest. See also color insert, Volume 1.

direction, they could travel a considerable distance. It is not uncommon for mangrove seedlings from Mexico, for instance, to be stranded and take root in Texas virtually across the length of the Gulf of Mexico. An even greater dispersal may explain the mangrove species *Rhizophora samoensis*, which is found only in Samoa and adjacent islands, at the opposite extremity of the Pacific from its presumed ancestor, the species *R. mangle* of Central America. The significance of dispersal ability for the geographical distribution of mangrove species is discussed in Section V.B.

II. MANGROVE ANIMALS: FAUNA OF TERRESTRIAL ORIGIN

Although mangrove roots are periodically immersed, the branches and leaves provide an environment little

FIGURE 3 Mangrove pneumatophores in Negombo Lagoon, Sri Lanka.

FIGURE 4 Mangrove propagules on a *Rhizophora* tree, Indus Delta, Pakistan. See also color insert, Volume 1.

different from that in adjacent terrestrial forests, with which they consequently share much of their fauna. Mangrove animals of terrestrial, rather than marine, origin include arthropods (particularly insects, but also spiders and myriapods), amphibians, reptiles, birds, and mammals. Virtually none are found exclusively in mangroves.

A. Insects

Anyone who has worked in mangroves can testify to the abundance of biting insects, particularly mosquitoes and "sand flies" or biting midges (Ceratopogonidae). Mosquito larvae develop in rot holes in mangrove trees, in semipermanent brackish pools, or in the water retained in crab burrows. In the latter case, one East African species, *Aëdes pembaensis*, ensures a suitable burrow environment for its larvae by laying its eggs directly onto the claws of the crab *Sesarma meinerti*. Prey of adult mosquitoes includes, apart from humans, a variety of mangrove mammals and birds, and in some cases it extends to fish.

Ants are often abundant in mangroves, including the aggressive nest-building weaver ants (*Oecophylla*) of the Indo-Pacific and leaf-cutter ants (*Atta*) of South America. Particularly complex relationships have evolved between ants, epiphytic "ant-house" plants, and mangrove trees (Fig. 5). Ant-house plants have bulbous stems (which may weigh several kilograms) honeycombed with passages inhabited by ants. One such plant, *Hydnophytum formicarium*, has specialized chambers in which ants deposit the remains of their prey, and from which the plant can absorb nutrients released by fungal action. The situation is further complicated by the presence of butterfly larvae (*Hypochrysops*) which feed on the ant-house plant and which are tended by the ants. The relationship therefore involves interac-

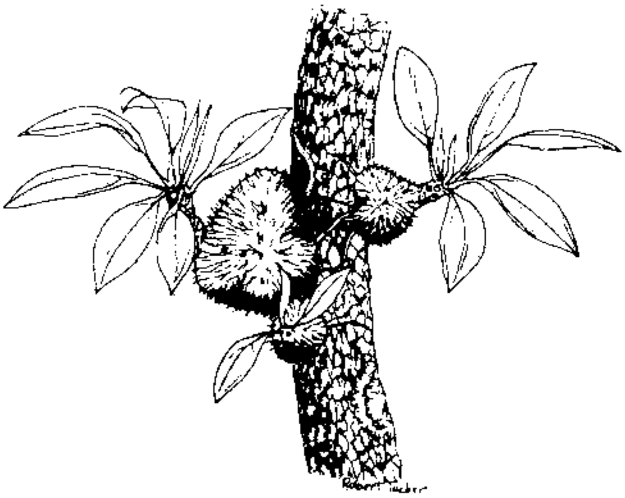


FIGURE 5 The "ant-house plant" *Myrmecodia*, epiphytic on a mangrove branch [reprinted from Saenger, P., Hegerl, E. J., and Davie, J. D. S. (1983). Global status of mangrove ecosystems. *Environmentalist* 3(Suppl. 3), 1–88, 1983, with permission of the International Union for the Conservation of Nature and Natural Resources].

tions between two plant species, two animal species, and one or more fungus.

Most mangrove ants are arboreal and essentially terrestrial animals. In many cases they nest outside the intertidal zone and forage in the mangal only at low tide. One Australian species, *Polyrachis sokolova*, is truly intertidal, retreating at high tide to nests within the mangrove mud. Nothing is actually known of its physiology: like other intertidal insects, it may retain a surface film of air and therefore avoid the need for any special adaptations to immersion or varying salinity.

Probably of greater ecological significance are the various plant-eating insects. Termites play a major role in disposing of dead wood. Some species construct nests of mud on tree trunks several meters above high-tide level, with access galleries snaking down the trunk to the aerial roots and upwards to the canopy.

The most important herbivores are those that eat mangrove leaves and seedlings, particularly the larvae of moths and beetles. Typically, only a small proportion of leaf production falls to herbivory. Sometime, however, it reached epidemic proportions. Individual trees in an otherwise healthy forest may be completely defoliated, and occasionally areas of many hectares are stripped of leaves. Canopy loss may result in the defoliated trees dying and being replaced by other species that are more tolerant of unshaded conditions. Insect herbivory therefore may alter mangrove community structure.

Other mangrove insects include the spectacular synchronously flashing fireflies of Malaysia (*Pteroptyx*), which occupy the mangrove *Sonneratia* for their displays, and numerous species of butterfly and moth. Hawkmoths, bees, and drosophilidid flies are among the species which are probably of importance in pollinating mangrove flowers.

B. Amphibia and Reptiles

Amphibia are rare in brackish or salt water, but one species, the crab-eating frog (*Rana cancrivora*), is common in mangrove habitats of Southeast Asia. Tadpoles survive well in salinity up to 50% that of seawater.

Reptiles are more abundant. Numerous species of snake forage within the mangal at low tide, including terrestrial or arboreal species but also some for which the mangal is their primary habitat. Mangrove snakes eat crabs (sometimes reciprocated), insects, and fish. In Southeast Asia, one of the most formidable mangrove predators is the monitor lizard (*Varanus indicus*), which may reach 1 m in length. Crocodiles, caimans, and alligators also occur in mangroves, although these are now rare in many areas due to human activities.

C. Birds

Birds are highly mobile. Many spend only part of their time in mangroves, migrating seasonally, daily, or tidally. Mangroves provide a feeding area, a nesting site, a refuge from the rising tide, or some combination of these. Waders probe for invertebrates in the mud of the mangal or adjacent mudflats. Kingfishers, egrets, and herons catch fish or invertebrates in the shallow water of mangrove creeks. Larger fish eaters, such as pelicans, ospreys, and cormorants, range further afield and may return to the mangal to roost or breed. In the Caribbean, roosts and nesting colonies of cattle egrets (*Bubulculus ibis*) and scarlet ibis (*Eudocinus ruber*) are so densely packed that the consequent enrichment of the soil with guano leads to significantly enhanced local growth of the mangrove trees.

Mangrove forests typically include numerous passerine species. Nectar feeders such as sunbirds in Malaysia, honeyeaters in Australia, and hummingbirds in South America move seasonally into mangroves and may be important pollinators. Insectivorous passerines specialize in hawking for insects in the canopy or, among low-lying vegetation, in picking insects off leaves or from bark crevices or from different species of tree. Broadly similar guilds of insectivorous birds, comprising different constituent species, seem to occur in different geographical regions.

Few of the species found within the mangal are mangrove specialists, and those which are restricted to mangroves in one part of the world may occupy different habitats elsewhere. One example is the cosmopolitan Great tit (*Parus major*), distributed from western Europe to China: only in Malaysia is it a mangrove species. The lack of mangrove specialists is probably due to the relative simplicity of the mangrove forest structure compared with typical tropical forest, allowing less scope for niche specialization. Another reason is probably the proximity of a pool of competing species in adjacent tropical rain forest. There are proportionally fewer mangrove specialists in New Guinea, where rain forest usually abuts mangrove habitats, than in Australia, where this juxtaposition is less common. Within Australia, there are few specialists in the mangroves of Queensland, which are extensive and contiguous with rain forest, than in northwestern Australia, where this is not the case. Most mangrove birds are probably using the habitat opportunistically.

D. Mammals

As with birds, many mammal species use the mangal opportunistically. These include small rodents, agoutis, wild pigs, antelopes, deer, and rhinoceroses; the Sundarbans of Bengal are the last major redoubt of the Bengal tiger (*Panthera tigris*). Domestic animals, such as camels and buffalo, are often a major element in the mangrove fauna. Otters may also be abundant, feeding on fish and crabs from the mangrove creeks.

Monkeys are common in mangroves. In Southeast Asia these include macaques (*Macaca*) which forage on the mud for crabs and mollusks. They also uproot large numbers of mangrove seedlings: Because these are seldom eaten or even greatly damaged, the purpose is not clear. Herbivorous monkeys are found in the forest canopy, including leaf monkeys (*Presbytis*) and, in the mangrove forests of Sarawak, the striking proboscis monkey (*Nasalis larvatus*). This is found only in mangroves and riverine forests, and it specializes in eating foliage, which is digested in an elaborate multichambered stomach with the aid of resident bacteria.

Bats are often abundant in mangroves. Resource partitioning in insectivorous bats parallels that of insectivorous birds, with species specializing in different zones of the mangrove vegetation and catching their prey with different flight techniques. A single bat may eat up to one-third of its body weight of insects each night: A 30-g bat might therefore consume 5000 insects nightly. The impact on the insect population of foraging bats must be considerable.

The exclusively Old World fruit bats often occur in mangrove forests in vast numbers: Roosts of an estimated 220,000 individuals have been recorded. Most fruit bats feed on nectar and fruit, and it is this which attracts many species into the mangal. In Malaysia, the long-tongued fruit bat *Macroglossus minimus* is an important pollinator of the mangrove *Sonneratia*: the long tongue is specialized for insertion into the *Sonneratia* flower, which carries large projecting stamens to deposit pollen onto the fur of the feeding bat. *Sonneratia* flowers last for only a single night, possibly because of the wear and tear resulting from visits by such a large pollinator. This species of bat is a true mangrove specialist, and in western Malaysia at least, it has not been recorded from other habitats. Mangrove specialization is possible only because the three species of *Sonneratia* in the area have different flowering patterns so that nectar is available throughout the year. Other fruit bats switch seasonally between mangrove and non-mangrove species.

III. FAUNA OF MARINE ORIGIN

One of the principal reasons for the high faunal diversity of mangrove ecosystems is their accessibility to occupation by organisms from both terrestrial and marine habitats. Of these, the marine invaders are the more numerous in terms of numbers and diversity of species. These include more or less sessile organisms settling on aerial roots and pneumatophores as well as more mobile species living on and under the mud. Many animal groups are represented in the mangal, the most conspicuous and ecologically most significant being teleost fish, crustacea, and mollusks. As with the land-derived mangrove fauna, the majority of species occur elsewhere and accumulate in mangroves because of the availability of food, shelter, or suitable substrate.

Considering mangrove communities at a scale of, for example, hectares, the diversity of such animals is often high. At smaller scale, however, the anoxic conditions caused by waterlogging, exacerbated by microbial decomposition of detritus, may greatly reduce both species diversity and abundance.

A. Root Communities

Mangrove roots and pneumatophores provide a hard substrate often covered with a rich and diverse growth of sponges, sea anemones, bryozoans, tunicates, barnacles, tubeworms, and mollusks as well as epiphytic algae. These in turn may attract a more mobile population of browsers or predators. The epibionts are mostly filter feeders, extracting organic particles suspended in the water, or predators of zooplankton, with no direct interaction with their mangrove host. A particularly thick growth, however, can adversely affect the host tree by occluding lenticels and restricting gas exchange with the underground roots. The relationship is sometimes mutually beneficial, as encrusting sponges may transfer nitrogenous nutrients to their host, and encrusting fauna can protect the root from attack by wood borers.

The labyrinthine aerenchyma tissue of the roots is easily penetrated by wood-burrowing organisms. The isopod crustacean *Sphaeroma* is a common root borer and may cause severe damage and even death. *Sphaeroma*-induced damage near the growing tip of a root may induce forking, with a resulting increase in the number of roots entering the soil: This may benefit the tree. The "shipworm" *Teredo* (which is in fact a mollusk) also bores dead roots and trunks extensively and plays a similar role to that of termites in disposing of woody debris. Like termites, *Teredo* relies on symbiotic microorganisms to digest the more intransigent components of wood.

B. Fish

Mangrove creeks and inlets are frequently occupied by abundant and diverse fish populations. In Southeast Asia, for instance, records of more than 100 species are by no means unusual. Many of these species spend only part of their time within the mangal, often moving to other habitats seasonally or at different stages of their life cycle. Mullets (*Liza*) eat significant amounts of mangrove detritus, such as shed leaves: most hunt small crustacea or other invertebrates. Some fish are permanent creek residents, commuting into the forest when it is submerged at high tide and foraging among the mangrove roots.

At low tide, Asian mangroves are occupied by mudskippers, which are relatives of the gobies (Fig. 6). As their name suggests, they skip across the exposed mud surface using their tails and leg-like pectoral fins, sometimes even climbing up aerial roots or pneumatophores. This amphibious life requires appropriate physiological adaptations, particularly in relation to respiration. Mudskippers are largely air-breathing, with gas exchange taking place not just across the gills but also at highly vascularized areas of the skin. Some store air within their burrows to enable aerial respiration even at high tide.

All mudskippers are probably to some extent omnivorous, but some are predominantly deposit feeders and others carnivores. Prey of the latter include crabs, insects, spiders, shrimps, and snails.

C. Crustacea

Mangrove habitats, particularly in the Indo-West Pacific, are dominated by crabs belonging to two families, Grapsidae and Ocypodidae. The former are predominantly herbivores or detritus feeders and the latter deposit feeders, extracting fine organic particles from mangrove mud. Predatory crabs, such as the formidable *Scylla*, may also be important components of the mangrove fauna. Shrimps (Penaeoidea) and mud lobsters (*Thalassina anomala*), and smaller crustacea such as amphipods and isopods, may also be significant as scavengers, in breaking down leaf litter, or as predators of smaller organisms.

1. Grapsid Crabs

Grapsid crabs of the subfamily Sesarminae, particularly of the genus *Sesarma*, are characteristic of mangroves, although a few species of this genus occur in other habitats (Fig. 7). More than 40 species of sesarmine have been reported from the mangroves of Malaysia alone, and many species, here and in other regions, undoubtedly remain to be described.

Sesarma are small (usually less than 3 cm in breadth) and inconspicuously colored. They are amphibious, re-

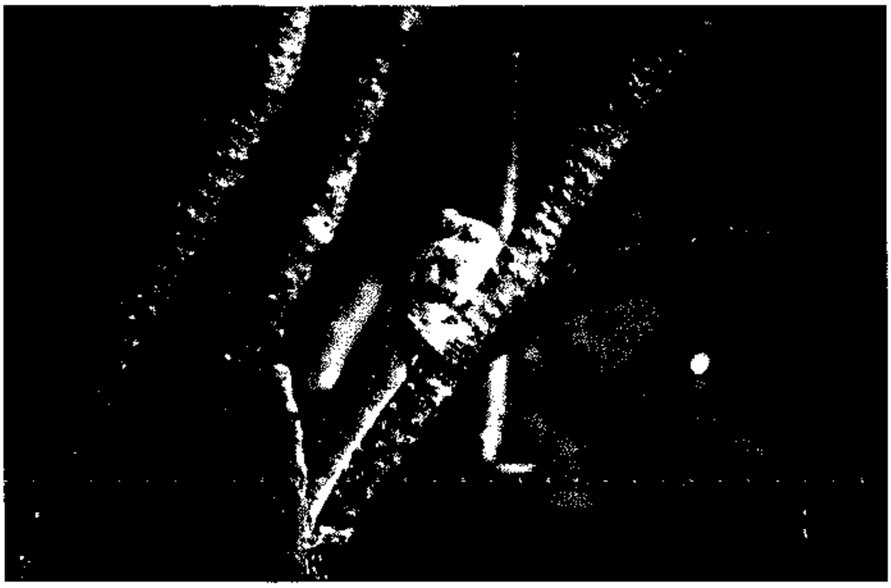


FIGURE 6 Mudskipper on an *Avicennia* pneumatophore (photograph courtesy of IIAR).

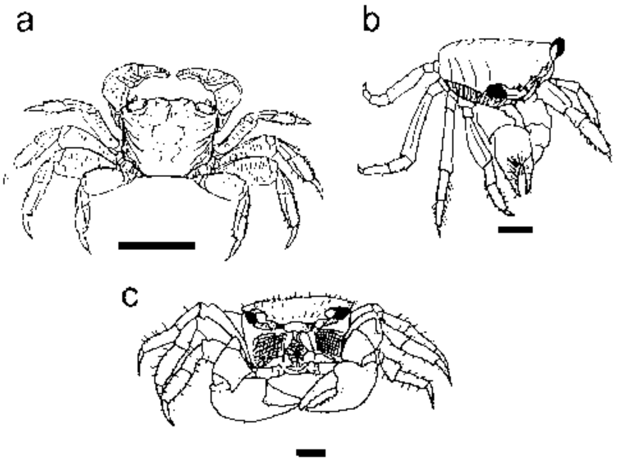


FIGURE 7 Mangrove sesarmine crabs: (a) *Parasesarma plicata*, (b) *Aratus pisonii*, and (c) *Neosarmatium smithi* [reprinted from Jones, D. A. (1984). Crabs of the mangal ecosystem. In *Hydrobiology of the Mangal* (F. D. Por and I. Dor, Eds.), pp. 89–109, with kind permission from Kluwer Academic Publishers and the author].

treating into burrows at high tide and foraging on the exposed mud at low tide. Respiration in air is achieved partly by recirculating water from the gill chambers over the carapace where it can be reoxygenated: evaporative cooling during this process also serves to reduce the dangers of high air temperature. Water loss can be offset by the acquisition of soil water through tufts of root-like hairs. Sesarmines are euryhaline, although differing degrees of salt tolerance probably contribute to the zonation of crab species along estuaries or with shore level.

In some cases, sesarmine crabs climb trees to feed on fresh leaves or buds. In East Africa, *Sesarma leptosoma* undertakes synchronized mass migrations twice daily from refuges among mangrove roots to forage on the tips of the branches of the trees. The virtually indistinguishable Caribbean species *Aratus pisonii* spends most of its time in trees, only rarely descending onto the mud.

Most sesarmines, however, subsist on fallen leaves or propagules. Mangrove leaves are often rich in tannins and other aversive materials, and several crab species have been shown to select leaves from the more palatable species of tree. Many leaves are collected as soon as they fall and cached in crab burrows. As decomposition proceeds, tannin levels decrease and nitrogen content increases through the accumulation of microbial biomass: storage therefore increases leaf palatability.

Much of the leaf material eaten is not assimilated but redeposited onto the mud as feces, available for microbial decomposition. It has been estimated that processing of leaf material by crabs increases the rate of breakdown of leaf litter 75-fold compared with the rate of decomposition under microbial action alone. Therefore, sesarmine crabs collectively play a very important role in facilitating energy flow through the mangrove ecosystem. By eating propagules, they also affect species distribution and community structure of mangrove trees (see Section V.C.1). However, there are geographical differences: in Southeast Asia and Australia, sesarmines are crucial in litter breakdown and selective removal of propagules, whereas in Florida and the Caribbean they are of lesser significance.

2. Ocypodid Crabs

Some crabs of the family Ocypodidae, such as the Central American hairy land crab *Ucides*, consume mangrove detritus. The majority are deposit feeders. Among these, the most conspicuous are the gaudily colored fiddler crabs (*Uca* spp.), widespread throughout the mangroves of the Old and New World (Fig. 8).

The common name derives from the one greatly enlarged claw of male fiddlers, which is used in courtship and in deterring rival males. The smaller claw of males and both claws of females are devoted to feeding. Mud is scraped into the buccal cavity in which, by a complicated process of flotation and manipulation by the mouthparts, fine organic particles are separated from the mineral components. The former is ingested and the latter deposited as a ball of sand, or "pseudofecal pellet." The process of separation may be quite selective. In some species, what is extracted consists almost entirely of microbial cells rather than, for example, fragmented leaf material. Others have subtly different extraction techniques and may specialize in the smaller meiofaunal animals. There may be as many as 60 fiddler crabs per square meter, resulting in 500 g of soil being processed daily. The toll on meiofauna is probably considerable, and the effects on soil texture and composition are profound.

3. Other Mangrove Crustacea

Other crabs found in mangroves are important predators. The most conspicuous is the mud crab *Scylla serrata* of the family of swimming crabs (Portunidae). *Scylla* reaches a carapace width of up to 20 cm, making it the largest invertebrate predator found in mangroves. Equally formidable predators are the mantis shrimps (Stomatopoda), which live in burrows in the mud and lacerate prey by rapidly shooting out their viciously spiked raptorial appendages. Other rarely seen burrowing crustaceans include pistol or snapping shrimps

FIGURE 8 Fiddler crab (*Uca*) in a Mozambique mangrove (photograph courtesy of D. Barnes). See also color insert, Volume 1.

(*Alpheus* spp.) and the mud lobster *Thalassina* (see Section III.C.4).

More general mangrove scavengers include hermit crabs, particularly *Clibanarius*, which forage on the mud surface at high tide. Shrimps may also be abundant in mangroves and mangrove creeks. Penaeid shrimps, which in at least some parts of the world depend heavily on mangroves for feeding and breeding, are an important commercial crop. The shrimp *Merguia* apparently lives only in mangroves and has the distinction of being the only semiterrestrial shrimp: it actually climbs trees. Only two species are known. One occurs in the Indo-West Pacific region, from Kenya to Indonesia, and the other occurs in Panama, Brazil, and Nigeria. Indo-West Pacific and Atlantic regions differ in the composition of their mangrove floras, and the separation of the two species of mangrove-associated shrimps may have occurred in parallel with the divergence of the mangroves themselves.

4. Crustacea as Ecosystem Engineers

All species have an impact on their environment, at the very least exchanging materials in the form of food, waste materials, and respiratory gases. Some species have effects beyond these simple transactions and alter the nature of their environment in ways that affect species other than their direct competitors, predators, or prey. Such species are often termed "ecosystem engineers."

In a mangrove ecosystem, the trees are the greatest engineers, influencing sedimentation rates and creating a physical environment. Crustacea also, in important ways, transform their surroundings. The topography of mangrove swamps in Southeast Asia is often visibly modified by mud lobsters (see section III.C.3). While processing mud, *Thalassina* throws up waste material from beneath the surface, forming mounds which may reach 2 m in height. These create patches of dry mud which provide habitats for other species, including the mangrove fern *Acrostichum*, fiddler crabs, and a variety of other burrowing crustacea and mollusks. Between the mounds the mud surface is lower, and more waterlogged, than it would be otherwise. Burrowing crabs also contour their environment, although less dramatically.

Much of the microbial activity of mangrove mud occurs in the surface layer, to a depth limited by the diffusion and exchange of gases with the atmosphere. As fiddler crabs process surface mud, they continually expose fresh material, facilitating microbial activity, while the active surface of the mud is increased in area by crab burrows. Burrowing activity also oxygenates the deeper soil and creates an underground labyrinth of interconnecting passages, through which significant underground water flow occurs. Experimental evidence suggests that crab activities significantly affect nutrient recycling and enhance growth of mangrove trees. Crustacea therefore alter the state of their environment in ways that significantly affect other species.

D. Mollusks

1. Bivalves

The most visible bivalve mollusks of mangroves are the oysters and mussels found attached to roots. Within the mud, however, there is often an abundant population of burrowing species. These, like the oysters and mussels, are largely filter feeders, extracting fine organic particles from suspension. A less typical group of bivalves are the shipworms of the family Teredinidae (see Section III.A), including the giant mangrove shipworm *Dicyathifer*, which may reach 2 m in length.

2. Snails

Gastropod snails are also generally abundant in mangroves. As with the crustacean fauna, these include herbivores, detritus and deposit feeders, and predators. Although a few species are uniquely found in mangroves, the majority of surface-living species also occur on open mudflats.

The principal predatory snails are species of *Thais*, found in mangroves worldwide. These cruise over mud and mangrove roots, feeding on barnacles or smaller gastropods. In the mangroves of Costa Rica, for example, *T. kiosquiformis* densities may reach more than 200 per square meter, and the species plays a major role in maintaining the function of mangroves by removing encrusting fauna from their roots.

Many gastropod species are deposit feeders, ranging in size from tiny and almost invisible species to the massive *Terebralia* and *Telescopium* of the Indo-Pacific region, which may reach a length of 10 cm (Fig. 9). One species, *Terebralia palustris*, feeds on small detritus particles when young, but on reaching a length of approximately 3 cm it switches to a diet of fallen leaves. The teeth on the radula (the ribbon-like tongue) of gastropods metamorphose appropriately to a form suitable for the altered diet. In Florida, snails are important consumers of mangrove seedlings, at some locations destroying nearly three-fourths of the seedling population. This is an interesting geographical contrast with other regions, such as Malaysia and Australia, where crabs fulfil this role (see Section III.C.1).

The most abundant snails on the mangrove trees are often species of *Littoraria,* close relatives of the periwinkles of temperate rocky shores. In Central America, on both sides of the Isthmus of Panama, the common species is *L. angulifera*. In the Indo-Pacific, this species is replaced by many others, which partition between them the slightly different habitats afforded by a tree. In Papua New Guinea, *L. scabra* prefers the bark of trees on the seaward side of a forest, *L. intermedia* prefers trees next to freshwater creeks, whereas the polymorphic species *L. pallescens* is found solely on leaves.

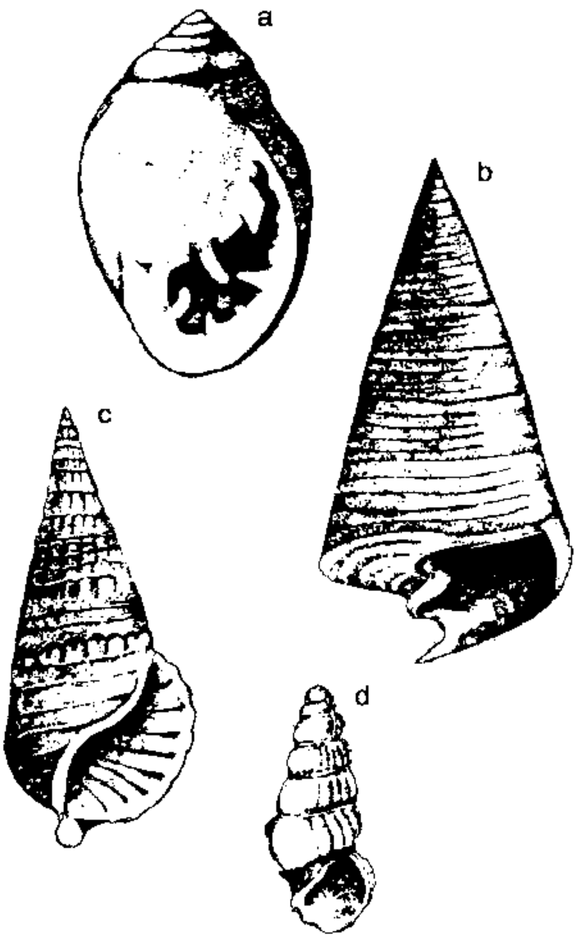


FIGURE 9 Gastropod mollusks typical of Indo-West Pacific mangroves: (A) *Pythia* (2.5 cm), (B) *Telescopium* (10 cm), (C) *Terebralia* (6 cm), and (D) *Cerithidea* (2 cm) (reprinted from *Advances in Marine Biology* 6, W. Macnae, A general account of the fauna and flora of mangrove swamps and forests in the Indo-West Pacific Region, 74–270, 1968, by permission of Academic Press).

E. Meiofauna

Within the mangrove mud lies a rich fauna virtually invisible to the naked eye—the meiofauna. Beneath an area of 10 cm^2 of mud there may be many thousands of individuals. Orders of magnitude smaller than the more conspicuous macrofaunal crabs and snails are meiofaunal herbivores, detritivores, and formidable predators, with food chains probably dependent on photosynthetic cyanobacteria ("blue-green algae") and heterotrophic bacteria. Meiofauna colonize fallen leaves, and the stages of leaf breakdown are accompanied by complex interactions and successional shifts in species composition and community structure which parallel, on a microscopic scale, the processes of macroecology.

The numbers of meiofaunal individuals are immense, and their diversity is astonishingly high. Not only are there many species but also the species show a higher level of taxonomic diversity. Among the macrofauna virtually all species belong to just three phyla: arthropods, mollusks, and chordates. The meiofauna from just one mangrove area in Australia, for example, yields turbellarian flatworms, nematodes, copepods, Ciliophora, Foraminifera, bivalve mollusks, oligochaete and polychaete annelids, hydrozoa, archiannelids, kinorhynchs, tardigrades, and gastrotrichs.

Very little is understood about the meiofauna of mangroves, their interactions, their functional significance in the ecosystem as a whole, and the relationship between the meiofaunal and macrofaunal worlds. Their small size belies their great importance.

IV. CONNECTIONS

The salient features of typical mangrove ecosystems are relatively high rates of primary productivity, much of the results of which enter decomposition pathways, either directly or after initial breakdown by leaf-eating crabs or mollusks. This is true of leaves and reproductive structures and, on a more protracted timescale, of the woody components of the trees. Particulate organic matter, either small leaf fragments or bacterial cells, is ingested by molluskan and crustacean deposit and filter feeders, enters meiofaunal food chains, or accumulates in the mud.

The ecosystem can be viewed physically as well as in terms of the flow of energy or matter. Mangrove trees supply hard surfaces on which other organisms settle, and they modify (as well as respond to) the physical environment by stabilizing the soil, facilitating accretion of mud, and retarding erosion. The environment is further modified by the physical activities of burrowing crustacea and other animals.

Mangrove ecosystems cannot be considered in isolation. They interact with adjacent habitats through the trapping of exogenous sediment or export of particulate or soluble organic matter or inorganic nutrients. Animals, by moving between mangroves and other habitats, also contribute to import and export of matter. Commercially important penaeid shrimps use mangroves as nursery areas so that shrimp catches many miles away may depend critically on mangrove productivity. Hard evidence for such connections between mangroves and other ecosystems, however, is sometimes elusive, and the strength of such linkages is almost impossible to quantify.

V. MANGROVE DIVERSITY

Mangrove diversity must be considered at a range of spatial scales, from global patterns of species richness to the pattern of distribution, at a particular location, at a scale of a few meters. In considering mangrove fauna, even smaller spatial scales become relevant. At all scales, diversity is affected by the past history of the area, by physical factors, and by biotic interactions, but the importance of each of these and the timescales over which they operate vary with scale.

A. Global Patterns

1. Latitudinal Range and Species Diversity

Mangroves are almost exclusively tropical or subtropical. This distribution is a reflection of a temperature limitation: The global distribution of mangroves correlates very closely with, for example, the winter position of the 20°C isotherm (Fig. 10). The number of mangrove species declines with increasing latitude, with the most northerly and southerly mangroves being species of *Avicennia*. In temperate regions, mangroves are replaced by salt marsh vegetation: plants which, like mangroves, are adapted to conditions of salinity and waterlogging but which do not carry the additional burden of being a tree or of producing large propagules.

2. Longitudinal Differences

Within their temperature and latitudinal constraints, mangroves show interesting patterns of species distribution. The principal biogeographic division is between the Indo-West Pacific (IWP) and Atlantic–Caribbean–east Pacific (ACEP) regions. These two regions have broadly similar areas of mangrove habitat, but the IWP has four times more genera and six times as many species of mangrove: 17 genera compared to 4, and 40 species compared to 7. It is apparent that none of the mangrove genera are very diverse, possibly because of a general limitation on species diversification in harsh intertidal conditions. Genera occurring in the IWP, however, are slightly more speciose than those of the ACEP: 2.35 compared with 1.75 species per genus.

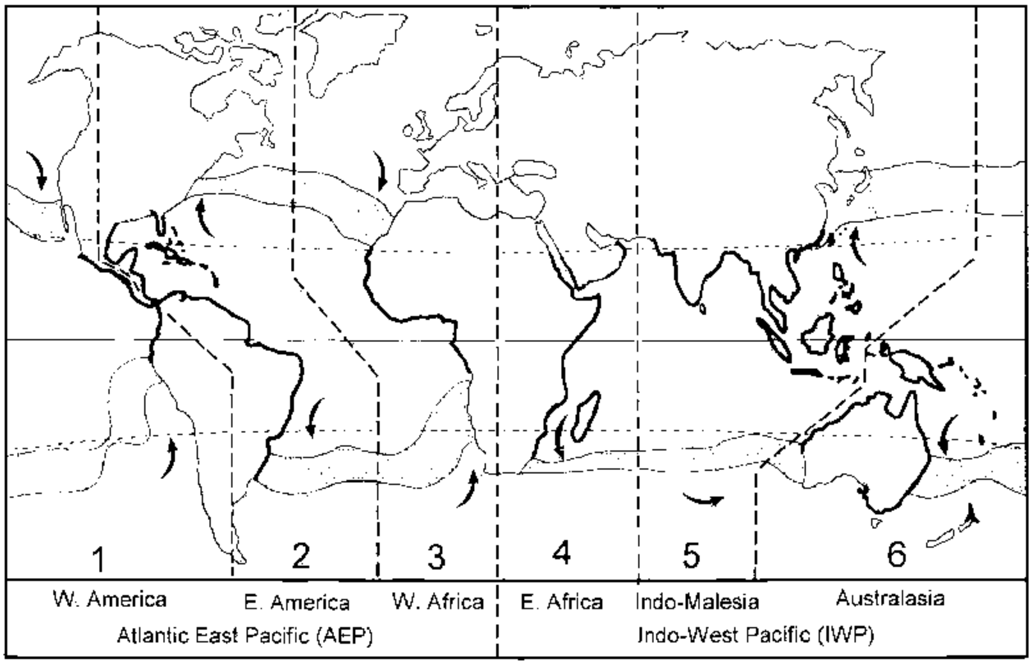


FIGURE 10 World distribution of mangroves in relation to 20°C isotherms [reprinted from Duke, N. C. (1992). Mangrove floristics and biogeography. In *Tropical Ecosystems* (A. I. Robertson and D. M. Alongi, Eds.), pp. 63–100, with permission of the American Geophysical Union and the author].

The differences between the IWP and ACEP regions are maintained by major barriers. The most obvious of these is the African continent (Fig. 10). Less obvious is the barrier represented by the central Pacific. This results principally from dispersal limitations rather than from the absence of suitable habitat. Suitable environments are present on many Pacific islands without natural mangrove populations, as shown by the success of the artificial introduction of mangrove species to Hawaii.

Further dispersal barriers, including the Isthmus of Panama, open ocean, and arid coasts unsuitable for mangrove occupation, divide the major regions into smaller subregions, each with a more or less distinctive mangrove flora (Fig. 11). Only one species occurs in all six subregions: the mangrove fern *Acrostichum aureum*. Two genera, *Avicennia* and *Rhizophora*, are common to both IWP and ACEP regions. All other genera are found exclusively in either the IWP or the ACEP, although the close similarity between *Laguncularia* (ACEP) and *Lumnitzera* (IWP) suggests a recent separation of these two genera.

The traditional explanation of mangrove species distribution is of a center of origin and of diversification in Southeast Asia, followed by dispersal restricted by physical barriers. This clearly makes little sense in relation to the current dispersal barriers. Fossil evidence of mangroves is widespread and reveals a much wider distribution during the Eocene and earlier epochs: Fossil *Nypa*, *Avicennia*, and *Rhizophora* pollen and other remains, for instance, have been identified in Eocene and Miocene deposits that now form part of North and South America, Europe, and North Africa as well as

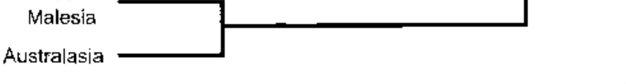


FIGURE 11 Comparison of the mangrove flora in six geographical subregions. Because of the recent closure of the Isthmus of Panama, the eastern Pacific and western Atlantic (including Caribbean) are most similar in species composition. Note also the separation between Atlantic–Caribbean–eastern Pacific (ACEP) and Indo-West Pacific (IWP) regions.

south and east Asia. At the time, these locations were connected by the Tethys Sea, continuous through what is now the Mediterranean and Indian Ocean.

Subsequently, this pantropical distribution was partitioned as a consequence of continental movements. Cosmopolitan genera such as *Avicennia* and *Rhizophora* were separated into regional populations by the approach of Africa to Asia 30–35 million years ago which closed the Tethys Sea, and separation of the sister genera *Laguncularia* and *Lumnitzera* followed the widening of the Atlantic barrier. The emergence of modern species ensued within the isolated subregions. Closure of the Isthmus of Panama was geologically very recent (a mere 2 or 3 million years ago) so that differences between eastern Pacific and Caribbean species are slight. One species (*Pelliciera rhizophorae*) is found on both sides of the Isthmus, presumably reflecting a separation into two populations too recently for allopatric speciation to have occurred.

An originally pantropical mangrove distribution was therefore partitioned into regions and subregions, with subsequent evolutionary divergence. Climatic conditions then eliminated mangrove species from areas such as southern Europe and the Mediterranean fringes. The current distribution pattern results from a combination of large-scale geographical factors and more regional climatic ones.

3. Diversity of Mangrove Fauna

It might be expected that faunal species diversity would follow a similar pattern to that of mangrove tree diversity both because the mangrove fauna has presumably been exposed to the same influences and because of a presumption that faunal diversity should respond to tree diversity.

The IWP region, richer in plant diversity than the ACEP, is also richer in species of mangrove-associated crustacea and mollusks (Table III). The reverse is true of other taxonomic groups, particularly those that form constituents of the root communities, such as sponges, coelenterates, and echinoderms. This may reflect regional differences in tidal range and availability of roots for settlement. For many groups, unfortunately, little comparable data are available and recorded species numbers reflect taxonomic interest and effort rather than the composition of actual species assemblages.

B. Regional Patterns of Diversity

Species diversity varies within regions in response to many different factors. The ACEP region, in addition to having fewer mangrove species in total, shows less differentiation between localities within the region, and all the species available in the geographical vicinity are likely to be represented at most locations.

Various factors may result in local variation in species diversity. Mangroves do not grow on rocky shores or in areas where fresh water is completely lacking (which is in part why all tropical shores are not dominated by mangroves). Stretches of inhospitable coastline therefore act as barriers which affect mangrove dispersal and geographical distribution. The arid shores of Soma-

TABLE III

Number of Species Recorded from Mangroves in Various Localities in the Regions Indicated[a]

	Atlantic–Caribbean–east Pacific	Indo-West Pacific		
Taxonomic group	Caribbean/W. Atlantic	East Africa	Indo-Malesia	Australia
Sponges/bryozoa	36	1	5	7
Coelenterata/ctenophora	42	12	3	6
Nonpolychaete worms	13	3	13	74
Polychaetes	33	72	11	35
Crustacea	87	163	229	128
Mollusks	124	117	211	145
Echinoderms	29	23	1	10
Ascidians	30	13		8
Fish	212	114	283	156
Reptiles	3		22	3
Birds	138		177	244
Mammals	5		36	7

[a] From Saenger *et al.*, 1983.

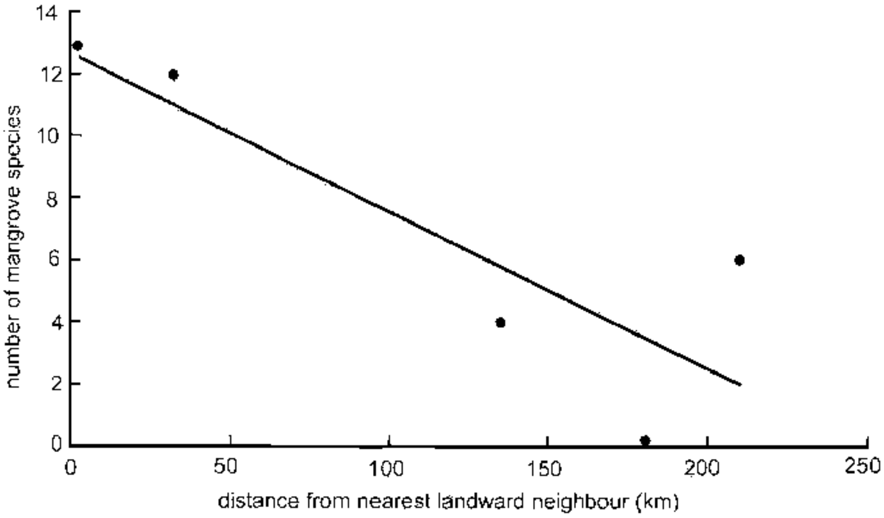


FIGURE 12 The number of species of mangrove occurring on West African islands in relation to their distance from the nearest landward neighbor [reprinted from Saenger, P., and Bellan, M. F. (1995). *The Mangrove Vegetation of the Atlantic Coast of Africa. A Review*, pp. 1–96, with permission of the Laboratoire d'Ecologie Terrestre de Toulouse].

lia, for example, result in the reduction in species number northwards so that *Avicennia marina* is virtually the only mangrove species found in the Red Sea. Separation of mangrove estuaries from each other by arid coastline, and regional-scale variation in physical variables, also affects the species distribution of mangroves around the Australian coasts.

Dispersal ability also affects species distributions within regions. The number of mangrove species on islands of the western Pacific shows clear attenuation with increasing distance from the species-rich areas of Australia and Papua New Guinea. Similarly, among islands off the West African coast there is a clear relationship between the number of mangrove species present and the distance from the nearest landward neighbor (Fig. 12). Species number also correlates with island size, with larger islands containing more species.

C. Local Variation in Species Distribution and Diversity

1. Tree Distribution

At a specific location, the distribution of mangrove species responds to physical variables in the environment. These often vary as gradients: in an estuarine mangal, for instance, salinity and the influence of tidal fluctuations tend to diminish with distance up the river. Sediment composition and nutrient dynamics also alter with distance from the open sea. Mangrove species respond differentially to such upriver/downriver gradients, resulting in zonation of species.

Similarly, in areas dominated by tide rather than river flow, tidal fluctuations establish gradients of physical variables, particularly in salinity and the extent of waterlogging of the soil. Again, mangrove species respond differentially to these physical variables and tend to form distinct zones. Where both river and tidal influences interact, the pattern of species distribution can be extremely complex.

In relation to salinity, species generally grow better at low salinity and differ more in the tolerance range than in their salinity optima. Low salinity, in consequence, tends to be associated with higher species diversity. At higher salinities, tolerance differences result in differing competitive success and translate into zonation of mangrove species along a salinity gradient, with species dominating zones at which they compete best, rather than those corresponding to salinity growth optima.

Although response to physical gradients suggests a gradual transition from one species to another as the determining physical variable gradually alters, this is often not the case. Mangrove species are frequently found in virtually monospecific stands or zones, with a more or less abrupt transition from one dominant species to another. This suggests that interactions between tree species, and mutual exclusion, may play a part in defining zone boundaries. Other physical variables, such as the degree of waterlogging and soil anoxia, nutrient availability, and biotic interactions between species, similarly affect species distribution within the mangal.

Superimposed on the sorting of species under the influence of physical variables are variations resulting from interactions with the mangrove fauna. Of the faunal influences, the most significant is the selective destruction of mangrove propagules by sesarmine crabs (see Section III.C.1). At least in Southeast Asia and Australia, this is a major factor determining mangrove species distribution. Mangrove animals respond to physical gradients of salinity and inundation regime. Sesarmine abundance is often greatest at midshore, and it is therefore here that mangrove propagules are most vulnerable. For reasons related to nutritional value and the levels of aversive tannins, *Avicennia* is generally the preferred food of sesarmines: hence at some locations the distribution of *Avicennia* in the upper and lower shore and their virtual absence from intermediate shore levels.

Random factors can also affect mangrove species distributions. If a gap is created in a mangrove forest because of the death of a tree, it is most rapidly filled by the species that are the best colonizers and best able to flourish in unshaded conditions. In Southeast Asia, the result is often an initial invasion of the mangrove fern *Acrostichum*. This may be succeeded by seedlings of *Bruguiera parviflora*. This species has relatively small and easily dispersed seedlings, whose growth is suppressed by the shade of an intact canopy. These in turn are replaced by slower growing shade-tolerant species such as *B. gymnorrhiza*. *Avicennia marina* is less tolerant of shade but is less likely to occupy a small gap because of propagule destruction by crabs. If, however, the gap is a large one, *Avicennia* is more likely to establish itself, probably because foraging crabs are vulnerable to predation in large open spaces. The distribution of species within a mangrove forest may therefore be patchy and reflect the stochastic nature of tree death and the subsequent successional history. On a larger scale, extensive death of trees by typhoons, by widespread defoliation by insect attack, or even by oil spills can have profound and long-lasting effects on species composition.

The structure of a mangrove forest is therefore in part explainable in terms of "patch dynamics"—of gaps appearing by chance and being filled by a changing assemblage of species differing in composition (at least for a time) from the surrounding forest. Eventually, something similar to the surrounding forest emerges. With a high incidence of gaps, a mangrove forest could be seen as a mosaic of patches of different successional age: if patches appear relatively rarely, the effect would be transient aberrations in an otherwise homogeneous, or consistently zoned, environment.

2. Distribution of Mangrove Animals

The species distribution of the mangrove fauna is less well understood since small, cryptic, and often mobile animal species are less easy to describe and analyze than large and immobile trees. A high level of taxonomic confusion compounds the problem. Nevertheless, it seems likely that the same general considerations apply. The distribution of mangrove crabs, for instance, forms zones related to shore level, salinity, and soil texture, whereas mollusks show zonation patterns in relation to shore level and to vertical position on the roots and trunks of mangrove trees.

The distribution of species of mangrove animals may also be related to patch size and the distance between neighboring patches, on a smaller spatial scale than applies to the distribution of mangrove species themselves, corresponding to the more limited dispersal ability of the species in question. This was demonstrated in the classical experiments of Simberloff on the terrestrial arthropod fauna (principally insects) of mangrove islets in the Caribbean. The species richness on a range of mangrove islets increased with the area of the islets and decreased with increasing distance from potential sources of fresh colonists. When the fauna of islets was completely eliminated with pesticides, recolonization soon established an equilibrium species richness similar to that before the elimination. In terms of the representation of different functional groups the previous situation was largely replicated, but the actual species comprising the new assemblages differed. Finally, artificially reducing the area of mangrove in experimental islets reduced species richness, showing that it was causally related to habitat area rather than to habitat diversity.

At an even smaller scale, individual mangrove roots can be regarded as "islands" of habitat suitable for epibiont settlement, surrounded by areas of unsuitable habitat. Here, too, the composition of root epibiont communities appears relatively stable in terms of functional groups. The actual species present are much more unpredictable and particularly affected by physical variables and by the supply of colonizing larvae. These factors are of different significance at different time and spatial scales.

Meiofaunal diversity has scarcely been investigated, although the same considerations apply as in the macrofaunal world. Variation in physical variables, species interactions, patchiness, dispersal, and the other factors relevant to larger organisms must also affect the meiofauna. To date, limited research interest (and the intrinsic difficulty of studying species interactions or measuring, e.g., nutrient gradients at a scale of millimeters) has restricted our knowledge of mangrove meiofaunal diversity and the factors which determine it.

D. Genetic Diversity of Mangroves

The advent of molecular genetic techniques has made it possible to study diversity at levels lower than the species. To date, few species have been studied, and clear general conclusions cannot be drawn. In some cases, such as the self-pollinating *Rhizophora mangle* of Florida and the Caribbean, populations appear to be genetically homogeneous, with slightly more genetic variation toward the northern extremes of the species' range. The extent of intraspecific genetic variation varies with the breeding structure of the population, with dioecious species showing much greater polymorphism. Genetic variation between populations is naturally greater than that within a population at a particular location, although West African mangroves have greater levels of genetic diversity than the same species in the Florida and Caribbean. This confirms the belief that western Atlantic mangroves derive from African populations rather than the reverse. As research proceeds, no doubt many such insights into the causes and consequences of intraspecific diversity will emerge.

VI. USES AND ABUSES OF MANGROVES

Mangroves are of interest not just to biologists. Their diversity and productivity makes them the source, directly and indirectly, of many products of use (and commercial importance) to humans.

Mangrove trees are exploited for timber for construction and firewood. This ranges from the casual collection of fallen wood to major charcoal industries based on the intensively managed mangroves of, for example, western peninsular Malaysia. Foliage may also be grazed directly or harvested for fodder for domestic animals. On a smaller scale, mangrove products are collected for a host of other purposes, including thatching houses, the manufacture of fish traps, for use in medicine, for tanning leather, and for use in various foods and drinks. Indirectly, mangrove productivity supports fisheries, both within the mangal and offshore. Less tangibly, mangroves can be of considerable importance in consolidating shorelines and limiting coastal erosion.

The significance of mangroves to humans varies greatly from place to place, but attempts have been made to achieve an overall economic valuation of the goods and services supplied. One recent estimate indicates that, on average, the annual value of a hectare of mangroves is approximately $10,000, resulting in a worldwide total contribution of $1,648,000,000.

An asset of this magnitude is worth conserving. Unfortunately, sustainable management of mangrove resources is the exception rather than the rule. In almost all parts of the world, mangroves are under pressure from irrigation schemes which divert rivers and prevent fresh water from reaching mangroves and from pollution, overexploitation, or deliberate clearance for construction or for the planting of alternative crops. One of the most destructive processes in many countries of Southeast Asia and Central America has been the clearance of mangroves for the construction of shrimp ponds—an attempt to increase the production of species dependent on mangroves while simultaneously reducing the primary production on which they depend. Not surprisingly, this has not been a success.

During the past few decades, loss of mangrove area in many countries has been dramatic. In the Philippines, for example, 60% of the mangrove area has disappeared, whereas in other countries such as Malaysia, Thailand, and Pakistan, annual losses are on the order of 1–3%. It may be, however, that the tide is turning. The virtual collapse of the shrimp industry in several countries and a greater awareness of the value of mangroves as a natural resource have focused attention on rational management strategies and on the possibility of reversing some of the damage. Much effort is now being put into replanting mangroves in abandoned shrimp ponds and the rehabilitation of denuded areas for coastal protection or in support of local fisheries as well as into developing suitable mangrove areas for ecotourism. The destruction of mangroves has largely been due to human activities: In the future their survival may also depend on mankind.

See Also the Following Articles

COASTAL BEACH ECOSYSTEMS • INTERTIDAL ECOSYSTEMS

Bibliography

Bandaranayake, W. M. (1998). Traditional and medicinal uses of mangroves. *Mangroves Salt Marshes* 2, 133–148.

Duke, N. C., Ball, M. C., and Ellison, J. C. (1998). Factors influencing biodiversity and distributional gradients in mangroves. *Global Ecol. Biogeogr. Lett.* 7, 27–47.

Field, C. (1995). *Journey amongst Mangroves.* International Society for Mangrove Ecosystems, Okinawa, Japan.

Hogarth, P. J. (1999). *The Biology of Mangroves.* Oxford Univ. Press, Oxford.

Hutchings, P., and Saenger, P. (1987). *Ecology of Mangroves.* Univ. of Queensland Press, Brisbane.

Macnae, W. (1968). A general account of the fauna and flora of mangrove swamps and forests in the Indo-West Pacific region. *Adv. Mar. Biol.* 6, 73–270.

Por, F. D., and Dor, I. (eds.) (1984). *Hydrobiology of the Mangal. The Ecosystem of the Mangrove Forests.* Junk, The Hague.

Robertson, A. I., and Alongi, D. M. (eds.) (1992). *Tropical Mangrove Ecosystems,* Coastal and Estuarine Studies 41. American Geophysical Union, Washington D.C.

Rützler, K., and Feller, I. C. (1996). Caribbean mangrove swamps. *Sci. Am.* 274(3), 94–99.

Spalding, M., Blasco, F., and Field, C. (eds.) (1997). *World Mangrove Atlas.* International Society for Mangrove Ecosystems, Okinawa, Japan.

Stafford-Deitsch, J. (1996). *Mangrove. The Forgotten Habitat.* Immel, London.

Tomlinson, P. B. (1986). *The Botany of Mangroves.* Cambridge Univ. Press, Cambridge, UK

ISBN 0-12-226868-7

90038

9 780122 268687

0-12-226533-7