Linking of Digital Images to Phylogenetic Data Matrices Using a Morphological Ontology

MARTÍN J. RAMÍREZ,[1] JONATHAN A. CODDINGTON,[2] WAYNE P. MADDISON,[3,4] PETER E. MIDFORD,[3]

LORENZO PRENDINI,[5] JEREMY MILLER,[6] CHARLES E. GRISWOLD,[6] GUSTAVO HORMIGA,[7] PETRA SIERWALD,[8]

NIKOLAJ SCHARFF,[9] SURESH P. BENJAMIN,[7] AND WARD C. WHEELER[5]

[1]*Museo Argentino de Ciencias Naturales "Bernardino Rivadavia"—CONICET. Avenida Angel Gallardo 470, C1405DJR, Buenos Aires, Argentina;*
E-mail: ramirez@macn.gov.ar
[2]*Entomology, Smithsonian Institution, PO Box 37012, NMNH E529, NHB-105, Washington, DC 20013-7012, USA*
Departments of [3]Zoology and [4]Botany, 6270 University Boulevard, University of British Columbia, Vancouver, BC V6T 1Z4, Canada
[5]*Division of Invertebrate Zoology, American Museum of Natural History, Central Park West at 79th Street, New York, New York 10024, USA*
[6]*Department of Entomology, California Academy of Sciences, 875 Howard Street, San Francisco, California 94103, USA*
[7]*Department of Biological Sciences, The George Washington University, Washington, DC 20052, USA*
[8]*Field Museum of Natural History, 1400 S Lake Shore Drive, Chicago, Illinois, USA*
[9]*Department of Entomology, Zoological Museum, University of Copenhagen, Universitetsparken 15, DK 2100 Copenhagen, Denmark*

Abstract.— Images are paramount in documentation of morphological data. Production and reproduction costs have traditionally limited how many illustrations taxonomy could afford to publish, and much comparative knowledge continues to be lost as generations turn over. Now digital images are cheaply produced and easily disseminated electronically but pose problems in maintenance, curation, sharing, and use, particularly in long-term data sets involving multiple collaborators and institutions. We propose an efficient linkage of images to phylogenetic data sets via an ontology of morphological terms; an underlying, fine-grained database of specimens, images, and associated metadata; fixation of the meaning of morphological terms (homolog names) by ostensive references to particular taxa; and formalization of images as standard views. The ontology provides the intellectual structure and fundamental design of the relationships and enables intelligent queries to populate phylogenetic data sets with images. The database itself documents primary morphological observations, their vouchers, and associated metadata, rather than the conventional data set cell, and thereby facilitates data maintenance despite character redefinition or specimen reidentification. It minimizes reexamination of specimens, loss of information or data quality, and echoes the data models of web-based repositories for images, specimens, and taxonomic names. Confusion and ambiguity in the meanings of technical morphological terms are reduced by ostensive definitions pointing to features in particular taxa, which may serve as reference for globally unique identifiers of characters. Finally, the concept of standard views (an image illustrating one or more homologs in a specific sex and life stage, in a specific orientation, using a specific device and preparation technique) enables efficient, dynamic linkage of images to the data set and automatic population of matrix cells with images independently of scoring decisions. [AToL; Araneae; bioinformatics; digital images; documentation; morphology; ontology; phylogenetics; spiders; systematics.]

Comparative biology seeks to synthesize all knowledge about the diversity of life on Earth. Over the last 250 years, taxonomists in particular have compiled large amounts of comparative information on taxa and species, especially on their morphology. However, acquisition of morphological data has always been difficult, and its full documentation and dissemination have been limited.

For example, the Biologia Centrali Americana (1879–1915) was among the largest such efforts ever published. It comprises 63 large, thick volumes and contains 1677 plates (900 colored) illustrating 18,587 subjects. It described 50,263 species, of which more than 19,000 were new (http://www.sil.si.edu/digitalcollections/bca/explore.cfm). The best documentation of morphology in such works is provided by illustrations, but this magnum opus illustrated only 37% of the species treated. Of those species, only a few aspects of morphology were illustrated, with an average of perhaps two illustrations per subject. The expense of publishing this work made it relatively inaccessible: only a few libraries contain a complete set of the Biologia Centrali Americana (none, ironically, in Central America). Eugène Simon (1848–1924) was the most prolific spider taxonomist but illustrated only 20% of his ca. 4600 descriptions. Tord T. T. Thorell (1830–1901) described ca. 1500 species but illustrated only 3 (0.002%). Perhaps 90% of post-1950 descriptions included at least one illustration, and virtually all post-1965 descriptions include illustrations, but the vast majority only of genitalia.

Through much of the 20th century, the cost in time and resources to produce illustrations (hand drawn, or film-photographed) and to publish them remained too high to permit copious use of images. Inability to document and disseminate morphological data, in turn, led to huge losses of comparative knowledge as generations turned over. Successive generations of specialists had to reevaluate that information. Information could not be efficiently preserved or disseminated.

The advent of digital imagery, personal computers, and information networks has the potential to eliminate this problem. Digital photomicrography, scanning electron microscopy, and technological advances for 3-D reconstruction of morphology, including confocal laser scanning microscopy (Klaus et al., 2003) and computer tomography (Wirkner and Richter, 2004), are accelerating the pace at which morphological characters are discovered, while a parallel "revolution" in cyber infrastructure is transforming the rate at which they can be documented and disseminated via the Internet (Agosti and Johnson, 2002; Bisby et al., 2002; Gewin, 2002; Godfray, 2002a, 2002b, Wheeler, 2003, 2004; Godfray and Knapp, 2004; Wilson, 2003, 2004; Thacker, 2003). Digital images

can be cheaply produced and disseminated electronically. Online repositories can capture their many attributes, such as the phylogenetic characters, specimens, and taxa they illustrate (e.g., Proszynski, 2003–2006; AntWeb, http://www.antweb.org/).

This new ability to produce, store, and disseminate many images poses a new challenge: how to organize and use these images efficiently for phylogenetic studies? Many web-accessible image databases exist, but systematics puts special demands on such systems. For instance, the repository should illustrate and justify the actual scores in matrix cells, as well as the concepts underlying each character and its states. It should offer queries of images based on homology hypotheses, and, ideally, facilitate discovery of more refined homology hypotheses. It should also be designed for collaboration and parallel workflows, integrating work of individual researchers and research groups into a common, publicly available repository that links images to phylogenetic studies.

This paper addresses the challenge of how to maintain, curate, share, and make efficient use of these collections of digital images. We specifically address how to link efficiently images to phylogenetic data sets and propose a solution based on an ontology of morphological terms. We aim for comprehensive, long-term, expandable and expanding morphological data sets spanning many specific analyses and multiple grant cycles and involving multiple collaborators and institutions; e.g., programs of the United States National Science Foundation such as Assembling the Tree of Life (AToL), Planetary Biodiversity Inventories (PBI), and Partnerships for Enhancing Expertise in Taxonomy (PEET; Rodman and Cody, 2003).

Although our own perspective is that comparative morphological data are vital for comprehensive and well-corroborated reconstruction of phylogeny, the need for database systems such as those we discuss does not depend on this perspective. If we are to document and explore comparative phenotypic data for any purposes, whether phylogeny reconstruction or interpretation of evolutionary patterns, efficient and phylogeny-aware image repositories will be needed.

BOTTLENECKS IN THE DOCUMENTATION, REPLICABILITY AND ACCUMULATION OF MORPHOLOGICAL DATA

Illustrations are essential at all stages of a phylogenetic study, from background examination of legacy data to exploration of new characters, and in principle should document all character states and cell scorings. At present, however, the process as a whole suffers from various limitations and bottlenecks.

Production.—The primary bottleneck in morphological data analysis is the production of the data themselves, even if the workflow is made more efficient (see below). Relatively few comparative morphologists are still active, and that number continues to decrease (Gaston and May, 1992; Systematics Agenda 2000, 1994). Training new morphologists ideally requires them to review all published morphological work in their field and to learn the often specialized and undocumented techniques required to acquire new data (Wheeler, 2004). Although digital imaging technology has mitigated the difficulties of producing and storing photographs, interpretative drawings are frequently mandatory. Even if computer generated, such drawings require much skill and time to produce. Also, although methods and protocols are becoming increasingly standardized, different labs or researchers may not image the same structure in the same way or from the same angle, leading to interpretative difficulties. Finally, an image is not interpreted data: the characters must be scored. Sifting through images of many taxa to formulate homology hypotheses and to achieve formal character state scorings is a laborious process with few technological aids.

Maintenance and continuity.—Maintenance and continuity of morphological data over intermediate and long-term time scales is another major bottleneck. Our costly data have often been ephemeral, inadequately documented, and thus lost to subsequent generations. Character or state definitions and scientific names of specimens inevitably change due to advances in knowledge, hypothesis testing, or corrections of error. Imaging of undescribed species is relatively common in groups where research on higher phylogeny outpaces descriptive taxonomy. Currently, such work generally results in a terse analytical publication including the phylogenetic data set, a list of specimens examined, verbal descriptions of characters and states, and perhaps a few dozen exemplar images to illustrate new or problematic characters. No mechanism is routinely applied to update these data.

Fixation of meaning for anatomical terms and characters.—Characters or states defined only verbally can be misinterpreted by other workers so that the meanings of terms drift and change from one work to the next. Because morphological terms are not unambiguously anchored to real examples, or "typified," comparative morphology still suffers from many of the same problems that faced taxonomic nomenclature prior to adoption of the name-bearing type system. The value of a stable taxonomic nomenclature is taken as a given, but the value of a stable character nomenclature is underappreciated. Because the systematics of large clades is of enduring scientific interest, presumably successive generations of comparative morphologists would find long-term maintenance a worthwhile investment if a mechanism to maintain character stability were available.

Publication.—Publication does not alleviate these bottlenecks. Probably no comparative morphological publication has ever included all the images on which it was based or those the author considered useful or relevant, absent constraints on publication. Most unpublished legacy data are lost over time. In order to use previously published concepts and discoveries, authors usually must reimage specimens. Traditional publication also does not easily accommodate the detailed metadata required to trace observations or images back to specimens in public collections. Terminals tend to be scored from multiple specimens (males, females, dissections,

specimens vouchering field notes and photos, etc.). If compiled in an appendix to a traditional publication, the complete list of individual source specimens and the specific observations they vouchered would be long, repetitive, and difficult to use, although online resources such as Proszynski's (2003–2006) diagnostic drawing atlas and AntWeb (http://www.antweb.org/) are a huge advance.

Analysis.—Analysis of phylogenetic data is not the bottleneck that it once was, due to new algorithms and parallel processing architectures (Goloboff, 1999; Nixon, 1999a; Janies and Wheeler, 2001; Ronquist and Huelsenbeck, 2003; Stamatakis et al., 2005). Nevertheless, the numbers of taxa that can effectively be included in morphological data sets is still limited by the bottlenecks discussed above.

The net effect of these bottlenecks is that many data must be reproduced from scratch for further analyses of the same organisms. Accumulation of reliable data with proper provenance and metadata is impeded. The same problems do not impede accumulation and synthesis of phylogenetic hypotheses expressed as trees, as the burgeoning field of supertree construction clearly demonstrates (e.g., Page, 2004a, 2004b).

SUPERMATRICES AND LEGACY DATA

We are instead concerned with the analogous problem of assembling supermatrices from legacy data sets. The AToL: Phylogeny of Spiders project (henceforth "Spider AToL") is a publicly funded multiperson and multi-institution endeavor to solve a large phylogeny problem, the relationships of all 111 families of spiders, a project that would require many individual lifetimes to complete (http://research.amnh.org/atol/files/). To summarize previous work and to maintain scholarly continuity with it, we fused all quantitative, or even semiquantitative, published matrices that treated three or more spider genera. These 67 data sets were produced by 30 different authors over 27 years and nominally comprised 1437 genera and 4395 characters (roughly 3600 genera of spiders are described; Platnick, 2006). If the same characters and states appeared in different matrices (with no conflict in scores for shared terminals), fusion was relatively straightforward, although shifts in character or state concepts from one study to another with no change in wording were undetectable. This operation resulted in 945 remaining genera and 3280 characters. More problematic were cases in which the terminals were the same conceptually (i.e., congeneric) but based on different exemplar species. On the one hand, fusing terminals would result in many polymorphic codings; on the other, retaining all terminals (Prendini, 2000, 2001) inflates the size and reduces the power of the matrix to summarize previous knowledge. Semantic variance in character description made probably identical characters (or character states) appear different (e.g., "carapace ornamentation" versus "carapace sculpture" versus "carapace texture"). Character states sometimes overlapped (e.g., "convex to oval" versus "oval to flat," or ranges of

meristic counts). Potential logical problems arose when authors coded multistate characters differently (e.g., as one versus two lines of data, or unordered versus ordered). The least problematic set were truly different characters that could not be fused in any way, although those introduced numerous missing entries. Identifying and organizing those characters was the goal of the exercise: to assemble all known, potentially informative, independent homology hypotheses in spiders and outgroups. Even though most source data sets had relatively few missing data, the resulting supermatrix, which contained more than 3 million cells, was 94% empty (see also Driskell et al., 2004).

Aside from operational and logical problems involved in their synthesis, legacy data usually lack adequate metadata. Voucher specimens, if they can be located, might have been taxonomically revised or otherwise reidentified. Potentially ambiguous characters, states, or cells may be imprecisely defined or insufficiently annotated. Standards in phylogenetic analysis and documentation have improved over the last 30 years but still vary greatly from one study to the next, which makes it difficult to judge the quality of legacy data (e.g., see Jenner, 2001). Uncritical recycling of legacy data and the homology hypotheses they represent is therefore inadvisable. Ideally, every cell in a morphological data matrix should derive from an investigator-credited observation, and nearly all should be photo-documented in order to minimize the chance that future workers will need to repeat the observation and to maximize longevity and value of the data.

PHYLOGENETIC DATA SETS AS ORGANIZERS OF THE COMMUNICATION AND PRODUCTION OF MORPHOLOGICAL DATA

Phylogenetic analysis of morphological data is now a fairly mature field. Modern comparative anatomy courses usually present anatomical terms in a cladistic context, often as character states mapped on trees. Standards for phylogenetic analysis of morphological data are clear and broadly applied across botanical and zoological domains, so that any well-trained systematist can produce many original observations and publish them in respected journals. Representation of anatomy as discrete characters and states, although controversial theoretically (e.g., Sattler, 1996), is now a standard way to summarize comparative data (e.g., Soltis et al., 2005; Brusca and Brusca, 2002).

Such lists of phylogenetic characters and states discipline data collection and structure the communication of results. For example, since the publication of the first quantitative analyses of the broad relationships of spiders (Coddington, 1990; Platnick et al., 1991; Griswold, 1993), subsequent authors (e.g., Hormiga, 1994a, 1994b; Silva Dávila, 2003; Schütt, 2003; Ramírez, 2000; Raven and Stumkat, 2005) have accepted, elaborated, and expanded on the initial character concepts. This growing corpus of explicit homology hypotheses increasingly guides the orderly examination of major

character systems such as somatic morphology, male and female genitalia, spinnerets, and behavior.

As homology hypotheses and commentary on them multiply, the need for scholarly documentation and synthesis grows increasingly acute. Platnick et al. (1991) and Griswold et al. (1998) published scanning electron micrographs (SEMs) documenting spinneret morphology in all the terminals of their analysis. Hormiga (1994b) and Scharff and Coddington (1997) provided illustrations for all morphological character states of araneoid spiders, and Griswold et al. (2005) published a collection of 1075 digital images documenting nearly all their character systems and scorings. A substantial proportion of these characters are now canonical hypotheses, and a parallel trend towards canonical images is clear, such as SEMs of spinneret spinning fields, trichobothria, tarsal organs, and ventral and retrolateral views of male copulatory organs or the standard diagnostic illustrations used to describe species. At the same time, falling costs in the production of illustrations caused by digital imaging technology has enabled the production and storage of far more illustrations than can ever be published on paper. The amount of image data documenting comparative biology has therefore increased explosively. Access to excellent collections from all continents and funding opportunities for large-scale, collaborative phylogenetic studies further fuel the increase.

SPECIMENS AS REFERENCE POINTS FOR PHYLOGENETIC DATABASES

The taxon-character data set cell in a cladistic analysis is usually considered the unit item (e.g., Nixon et al., 2001; Dettai et al., 2004), and it is displayed as such by matrix editors such as MacClade (Maddison and Maddison, 2000), Winclada (Nixon, 1999b), or Nexus Data Editor (Page, 2001). Mesquite (Maddison and Maddison, 2006) goes further by allowing multiple author-dated annotations to a single cell. The preceding discussion shows that many problems in comparative data management result from inadequate links to original sources. The data set cell is actually less fundamental than the specimens, images, and observations used to generate the data.

A data set cell is based on observations of specimens. How do we record a reference to these specimens? Specimen databases are increasingly standardized and accessible over the Internet; e.g., from the Global Biodiversity Inventory Facility data portal (www.gbif.net), which is moving towards the use of unique and stable identifiers (GUIDs, Globally Unique Identifiers for Biodiversity Informatics; see http://wiki.gbif.org/guidwiki/) for specimens in collections. When such resources are in place, linking images or cell scorings to unique specimen identifiers ought to be straightforward. Both GBIF and the Taxonomic Databases Working Group (TDWG, http://www.tdwg.org/) are converging towards the adoption of Life Sciences Identifiers (LSIDs, http://lsid.sourceforge.net/) as GUIDs for specimens and images, which can be resolved to deliver metadata in standard formats, such as RDF (see Shadbolt et al., 2006).

Observations of character states based on such images are then indirectly linked to specimens via unique identifiers (and additional fields for author, date, and other observation metadata), thus producing a specimen-based phylogenetic database. Such a database would be "upstream" of, and more fine-grained than, the conventional taxon-character matrix because more than one observation or image can substantiate a cell. For example, Daicz and Pol (personal communication) are developing a data set editor based on specimens in which cell values are generated on the fly as the union of observations from more than one specimen.

Specimen-based, rather than cell-based, databases better accommodate updates, such as corrections in specimen identification and taxonomic status and the fusion or splitting of terminals, characters, and character states. As knowledge progresses, characters are often redefined. The limits and number of states fluctuate over time, even within the same study. A character originally proposed as "aggregate silk gland spigots: (0) absent; (1) present" might be scored for many terminals before it becomes apparent that some clades are sexually dimorphic. Characters for each sex are then required. If some of the cells were initially scored indirectly by inferences from silk samples (viscid droplets on silk samples indicate the presence of aggregate gland spigots), but later it was discovered that some males steal female webs, the cells scored from "male" silk samples must be scored again as missing entries. A database based on specimens and their images with appropriate metadata makes such adjustments easier, without reexamining specimens, and more importantly, without loss of information or data quality. The use of resolvable GUIDs serving machine-readable metadata will allow automation of many of these operations.

MorphoBank (http://www.morphobank.org/) and MorphBank (http://www.morphbank.net/) both emphasize the importance of specimen-based repositories. GenBank now incorporates fields for specimen data, following the Barcode of Life Initiative (http://www.barcodinglife.org/).

STANDARD VIEWS FOR EFFICIENT DATA COLLECTION

During data collection, a "longitudinal" pass (scoring one terminal for all characters) is usually fast and efficient because few specimens need to be prepared or manipulated. However, a longitudinal pass presumes stability of all character systems and complete familiarity with them, knowledge that typically characterizes the middle or end, rather than the beginning of a project. A "transverse" pass (scoring one character for all terminals), on the other hand, requires the preparation and manipulation of many specimens and is in general inefficient. Many experimental characters will be discarded or redefined as the study progresses, requiring multiple transverse passes. Storing and retrieving primary observations, especially images, can make transverse passes faster, because specimens are handled only once. If images and associated metadata were attached to cells *before*

scoring commenced, the work would be more efficient, as well as better documented, more effectively conserved, and more easily communicated.

Attaching images to cells before scoring is not a trivial task. The Spider AToL project envisages a data set of 500 terminals by 1000 to 2000 characters, implying 500,000 to 1,000,000 cells, most of which ideally would be photo-documented (thus megabytes of data per cell). Although some characters may not require image documentation, that many manual linkages is still impractical. Static links also fail to address the problems of durability and maintenance described above for the extended use of legacy data, because as characters are reviewed, the previously associated images must be reviewed as well. Even if manual linkage of images to cells in a conventional phylogenetic data editor sufficed for documentation, it could not, for example, retrieve just those images relevant to a particular character before the cells are thoroughly curated. Formulation of new character hypotheses requires examination of relevant images across many terminals. Proper design of the database and interfaces makes such tasks more efficient.

Efficient linking of images to cells before scoring can be achieved via standard views. Because the images that document specimens and observations in phylogenetic studies are increasingly stereotyped both in content and orientation, most characters link naturally to standardized views. We define "standard view" as (1) a homology term (body region, behavioral unit); (2) a sex and stage (e.g., adult female); (3) a specific orientation (e.g., dorsal); and (4) a specific imaging device and preparation technique (e.g., SEM, trypsin digest; for a more complete model, see Blanco et al., 2006:66). One standard view can document several characters, such as a SEM of the female cheliceral promargin that documents characters of the fang, setae, and teeth. Once a character is associated with one or more standard views, linking its cells to images is also straightforward because every image is also associated with a taxon. The documentation defining the standard views simplifies the imaging process, which can then be more easily delegated to someone who is not an expert in the taxonomic group. Recording specimen and standard view identifiers at the moment of production of images adds important metadata to the images (provenance, homology term, orientation, device) at very low cost. The documentation of standard views for the Spiders AToL project can be consulted in http://research.amnh.org/atol/files/.

Automatically populating cells with images via standard views compartmentalizes the workflow. Image production and addition of metadata can be separated spatially and temporally from scoring. If new images are obtained after a cell is already scored, their standard view assignments automatically allocate them to the relevant cells and researchers are easily notified that new images require review. The same occurs when newly defined standard views are added to the project workflow. Cells can still be commented with ad hoc, labeled, annotated images. If characters are fused or subdivided, the cells they formerly referenced and their linked images

are automatically updated. For publication, archiving, or similar purposes, the dynamic links can be converted to hard-coded links between cells and the unique identifiers of images, thus producing a snapshot of the data set at a given time. Programs such as Mesquite can store such static links as cell comments with author, date, and some explanatory text. However, standard views enable the vast majority of cells to be populated automatically so that for an ongoing project, manual links become the exception rather than the rule.

Linking Images to Cells

Specimens as reference points and standard views lead to a simple protocol for linking images to cells. Each image references the specimen from which it was taken and the standard view that it depicts. To find the images that pertain to a particular cell (terminal taxon × character), a database or client program cross-references the specimens of that terminal with the standard views depicting that character and retrieves the relevant images.

A preliminary implementation of such a scheme is available in a beta version of the SILK package of modules for Mesquite (Maddison and Ramírez, 2006; see below, Figs. 3, 4). Currently, the SILK package takes on the burden of finding the terminal to specimen links and character to standard view links, but it would be possible to put this burden on a database, thus permitting the client program Mesquite to make the simple query "What are all the images for this terminal and this character?"

Several issues will complicate implementations of databases to store and client programs to access images in this way. For instance, a user may change names of terminal taxa and character after deposition of the images into the database. This requires the use of unique and stable identifiers for taxa and characters to aid in relocating images; if unique identifiers for taxa are well discussed today (e.g., LSIDs), identifiers for characters and character states are more problematic (see below). Also, not all characters in all taxa will be adequately illustrated through standard views. Ad hoc attachment of images to cells will be needed to deal with special cases.

Character and Character State Typification

The preceding discussion argues that images can clarify the meanings of cells in phylogenetic matrices. Names of taxa in matrices are fixed nomenclaturally by type specimens. However, names of homologues (characters and character states) are not currently "fixed" by any sort of typification procedure, and, not coincidentally, their meanings are subject to eternal debate. If the meaning of such terms is free to vary, it will not be useful, for example, to assign global unique identifiers to characters and states. Nomenclatural holotypes fix species names essentially as ostensive definitions or labels that point to one, unique object. A holotype does not "define" a species scientifically, it merely provides the objective reference for a name to enable accurate communication.

Although the confusion engendered by the lack of objective references for names of homologues is analogous to that which plagued nomenclature prior to the type system, it would be unwise to fix homologue definitions by literally designating particular specimens as types or to ape the rules of taxonomic nomenclature. For one thing, such specimens would require special status in museum collections, and no additional resources exist to curate them. For another, precise characterization of homologues often requires destructive sampling, leading to a paradox in which no specimen could be both pristine and proved to have the feature. The analogy to holotypes in taxonomy should therefore not be taken too literally. Taxonomic nomenclature may need elaborate, legalistic rules, but clarifying the meaning of a character or character state often simply requires an unambiguous (ostensively referenced) image. Homologue definitions can be fixed by designating a particular structure or condition in a particular species as the standard of reference or "type" (Hormiga, 1994a:5; Scharff and Coddington, 1997:371). Because species names are already typified, homologues would be no less objectively defined in the ultimate sense. For example, the male spider genitalic sclerite "median apophysis" (perennially debated; Coddington, 1990) could be defined as that particular structure in *Araneus diadematus* (Clerck, 1757), and images of that sclerite in any *A. diadematus* male would for all practical purposes fix the definition of the homologue. Alternative interpretations of the same character could also be accommodated, e.g., "median apophysis *sensu* Lehtinen 1967."

In this schema, such "type images" attach to character or character state names rather than data cells in phylogenetic matrices and therefore round out the fixation of all matrix elements (characters, taxa, and cells). Images attached to data cells then become hypothetically (or subjectively) homologous to type images. The latter, therefore, would be the same image (or images of the same structure in the same species) in all matrices referencing that feature.

Any arbitrary system for the fixation of names requires procedures to do so but also the social consensus to abide by them. As an experimental implementation, the Spider AToL intends to attach exemplar images to all character states whose interpretation might be ambiguous. The images can be displayed in a panel besides the cell images (see Fig. 4). The typification of character state concepts by such exemplar images may at first be provisional but should become progressively more stable after cycles of character study. At some point the typification should reflect stable consensus and would be effectively permanent, thus documenting character and state concepts, and may serve as reference for stable, unique identifiers.

STRUCTURING COMPARATIVE DATA
IN A HIERARCHY OF HOMOLOGUES

The standard view identifiers indicate what can actually be *seen* in the image and are easily mapped to an anatomical atlas of the taxon under study. Although their anatomical relations are not strictly relevant if standard views function simply as a flat data table to retrieve particular images from a large collection, the latter approach has limitations. First, significant numbers of images are not standard in various ways; e.g., an unusual angle, or a close-up rather than full frame, or perhaps produced from a different device or preparation technique. Legacy images are frequently nonstandard. Assigning such images to the closest standard view relaxes the rigor of standard views, which is undesirable. These nonstandard images still have to find their way to the data set cells and to character system specialists.

Second, as the number of standard views grows, managing views and curating the image collection becomes problematic. Dorsal, prolateral, ventral, and retrolateral views of the seven articles on all four legs on one side of a female spider yield 112 views. If the 376 standard views currently identified by the Spider AToL project were simply a flat list, routine tasks such as assigning the correct standard view identifier to an image would require perusing the entire list. To link a character of the anterior lateral spinneret spinning field to a view, one wants to see just the short list of views illustrating the spinnerets, or even better, the anterior lateral spinneret.

Both problems can be alleviated by grouping the anatomical terms and the corresponding standard views in a hierarchy of homologues according to part-whole relationships like titles and subtitles in an anatomical atlas: the anterior lateral spinneret spinning field is part of the anterior lateral spinneret, which is part of the abdomen (Fig. 1). Once the standard views are organized hierarchically, and the nonstandard images are linked to anatomical terms, they become accessible to automatic queries. The same hierarchy can organize the characters so that managing thousands of characters is much easier. Once the images and characters are structured according to a common hierarchy, the linking of images to characters, and the administration of the whole system becomes conceptually transparent.

AN ONTOLOGY OF HOMOLOGUES

The hierarchical organization of terms for homologous parts, structured by a part-whole relationship, constitutes a type of ontology (technically a mereology). Ontologies are an increasingly popular combination of a controlled vocabulary of terms with a relatively small set of logically defined relationships (Smith, 2004a, 2005; Trelease, 2006). Most biological ontologies include relationships for subsumption ("*is_a*") and part-whole ("*part of*"). Other examples of anatomical ontologies include the Foundational Model of [Human] Anatomy (Rosse and Mejino, 2003), the model organism anatomy ontologies for *Drosophila*, mouse, and zebrafish and the taxon-wide ontologies of anatomy of plants and fungi, all available from the OBO repository (http://obo.sourceforge.net/).

A well-constructed ontology is both logically consistent and accurately models the reality of its subject area (Smith, 2004a). Accurate modeling requires appropriate scoping of the subject area (e.g., considerations of
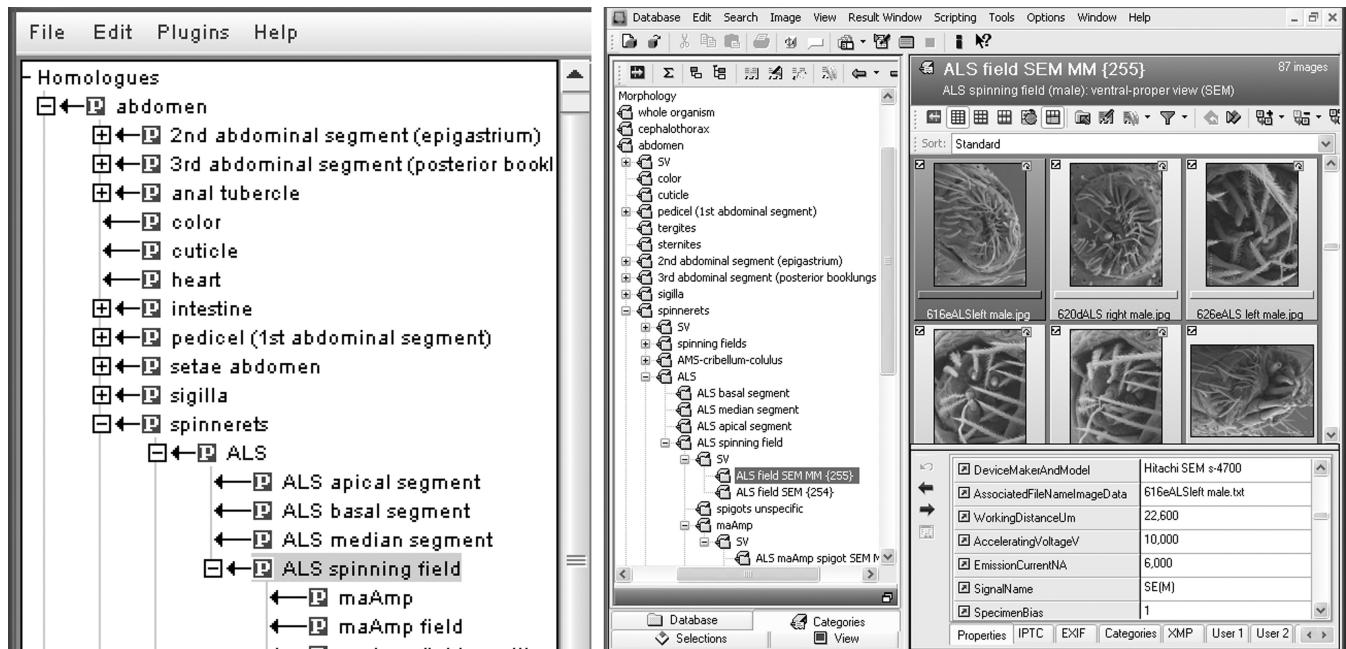
FIGURE 1. Left: Spider ontology of concepts of comparative biology as displayed by OBO-Edit (Day-Richter, 2001–2006). Right: Images and standard views linked to the ontology of homology areas as displayed by IMatch (Westpal, 2006).
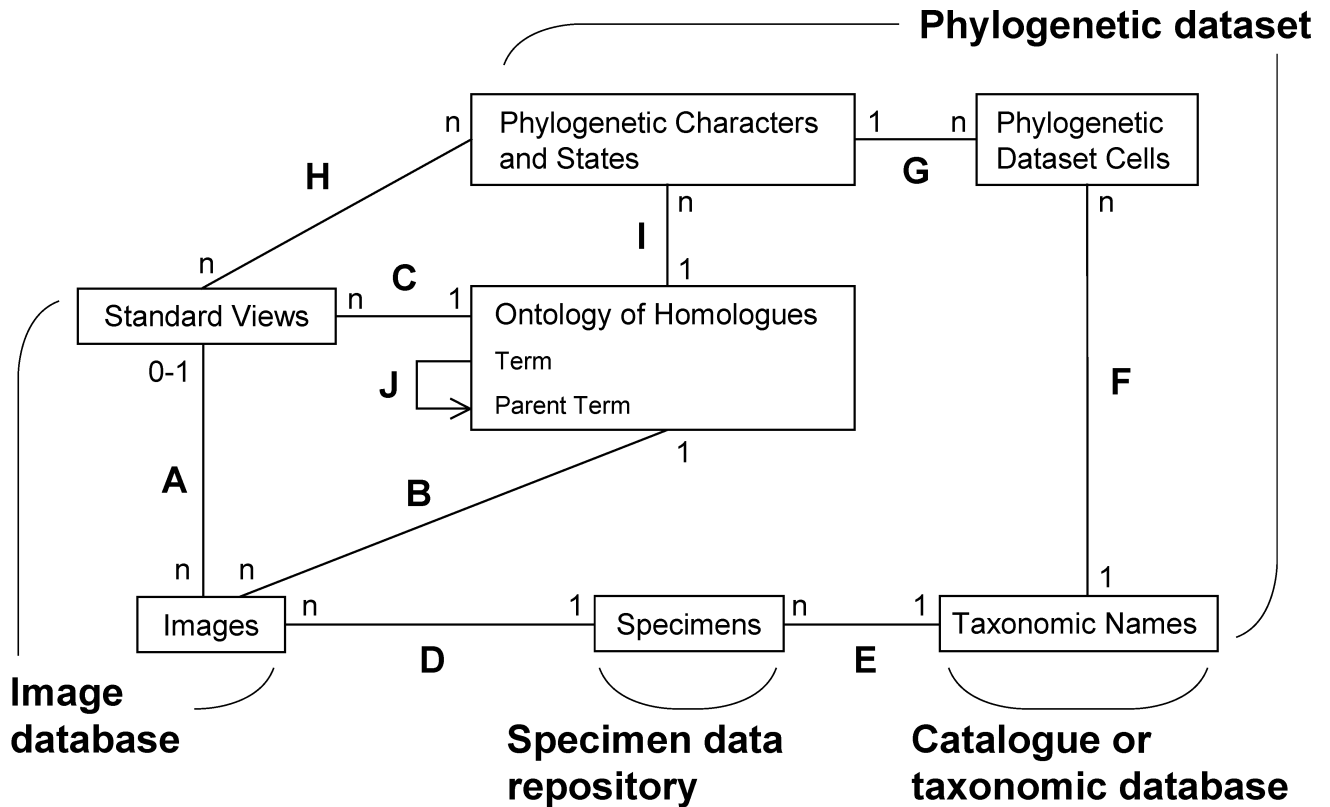
development, homology, or spatial proximity). Logical consistency requires that relationships be rigorously defined (e.g., Smith et al., 2005) and that term hierarchies and other asserted relationships between terms be consistent with those definitions. Consistency-checking tools such as OntoClean (Guarino and Welty, 2004) free biologists to focus on correctly modeling the domain. When properly constructed, ontologies facilitate communication among both humans and machines. Well-defined ontologies are particularly useful for applications involving machine reasoning and can increase confidence in software processing of massive amounts of data. By using unique identifiers for terms, the relationships can be adjusted without altering the underlying data.

The OBO format is an attractive platform for the construction of ontologies (Open Biological Ontologies; http://obo.sourceforge.net/). OBO format is used for numerous other biological ontologies, including the anatomy ontologies of model organisms mentioned above. The ability to learn from the experiences of these other anatomy projects and the availability of several supporting ontologies for relationships (Smith et al., 2005; http://obo.sourceforge.net/relationship/relationship.obo) and phenotype attributes (pato.obo at http://obo.sourceforge.net/) and tools such as OBO-Edit (Day-Richter, 2001-2006) made the OBO format an attractive choice for constructing our ontology. The OBO relationship collection includes most of the relationships necessary for modeling anatomy and other concepts useful in morphology. These include spatial relationships ("*located_in*", *adjacent_to*"), as well as temporal ("*transformation_of*," "*derived_from*") and those for describing

events and behaviors ("*has_participant*," "*has_agent*"). Further relations, which are not currently part of the OBO relationship collection, may be defined in collaboration with other large scale phylogenetic projects and submitted to the maintainers of OBO.

The spider anatomy ontology used in the Spider ATOL project is a taxon-wide ontology designed to accommodate the morphological, developmental, and behavioral characters used in higher level systematics (Fig. 1). At this moment, the working version includes only "*part_of*" relationships, which accommodates most of the homology terms used in phylogenetic characters. In a subsequent stage we will incorporate "*is_a*" relationships for serial and modular homology (= homonomy; e.g., leg IV *is_a* leg; trichobothria *is_a* seta).

All standard views and characters are assigned to terms defined in the ontology, and all ontological terms will be given explicit textual definitions and synonyms and linked to each other by subsumption, part-whole, and other logical relations. Because the logical relations required by ontologies are rather deeply connected to those required by programmers, the better an ontology meets ontological criteria, the more types of queries it will reliably be able to answer. Nonstandard images are also assigned to ontological terms but not to a standard view. In the example above, nonstandard views of the anterior lateral spinneret spinning field would simply be assigned to "ALS spinning field *part_of* ALS *part_of* spinnerets *part_of* abdomen." The ontology is also used to organize and segregate new images for review by curators, who may or may not assign them to standard views as appropriate. Using an ontology to structure the image database efficiently compartmentalizes and distributes

FIGURE 2. Schematic relations and main tables linking images to phylogenetic data set cells. The ontology of homologous anatomical terms is the central piece structuring image data and linking with the data set ($n = 0$ to many).

image-related work according to body regions or areas of expertise and manages characters similarly. The ontology, in fact, is the central organizing principle of this data schema (Fig. 2).

Figure 2 summarizes the main tables and relationships for the images, specimens, and phylogenetic data. The anatomical ontology is the central element that organizes the links between the phylogenetic data set and the images. We expect that in the near future the elements in this design could be distributed and maintained independently over the web, once GUIDs and reliable servers and interfaces are in place. For example, the phylogenetic data set could be hosted in MorphoBank, the image database in MorphBank, the specimen data accessed through the GBIF portal, the taxonomic names through the Taxonomic Search Engine (Page, 2005), and the ontology in obo.sourceforge.net. Each of these initiatives could provide the unique identifiers for each element and serve its associated metadata, and front ends like Mesquite could retrieve data items and infer relationships dynamically and transparently.

The SILK package of Mesquite uses simple tables (Fig. 3) derived from the ontology to display the images

in each cell. Whenever a character is added, it is sufficient to enter in the tables the identifier of the corresponding standard view, or in its absence, the identifier of the anatomical region, and the relevant images will appear in the cells (Fig. 4).

ONTOLOGY AS A RESEARCH TOOL

Retrieving a small set of highly relevant images versus a larger set with more images of the same body region are different tasks. The former can be obtained with a query based on the link between a character and a standard view, and the latter with a query based on the link between a character and an anatomical region in the ontology. The former suffices for fast scoring of a stable data set and for thoroughly imaged characters, but exploratory work requires the latter, perhaps all images containing that anatomical region, whether or not they are standard views. The image displaying the required structure in detail may be missing, but other, lower magnification images displaying the homologue may suffice. The position of features such as the tracheal spiracle can vary substantially between taxa and thus between

Character state image typification

CharID	CharName*	State	Comment	ImageFileName	LocalPath	RemotePath	Taxonomic Name*
1461	Cribellar spigots	0	...	03043004.jpg	...	...	Amaurobius similis
1462	Cribellar spigots	1	...	613gCribellum_female.jpg	...	...	Acanthoctenus cf spinipes

Matrix Cells

Terminal in Matrix	CharID	State
Amaurobius	1461	0
Amaurobius	1689	0
Amaurobius	3084	1
Ambicodamus	1461	-
Ambicodamus	1689	1

Taxa to terminals

Taxonomic Name	Terminal in Matrix
Amaurobius fenestralis	Amaurobius
Amaurobius similis	Amaurobius
Ambicodamus marae	Ambicodamus
Ammoxenus amphalodes	Ammoxenus
Ammoxenus coccineus	Ammoxenus

Characters to morphology

CharID		MorphologyID	
1461	(Cribellar spigots, 0: uniformly distributed 1: clumped)	SV253	(Cribellar spigots SEM, female)
1689	(PMS Cylindrical gland spigots, 0: absent, 1: present)	SV266	(PMS spinning field SEM, female)
1689	" " " " " " " "	SV267	(PMS spinning field SEM, male)
3084	(Number of heart ostia pairs, 0: four, 1: three, 2: two)	SR130	(Heart)

Image data

Taxonomic Name	MorphologyID	Comment	ImageFileName	LocalPath	RemotePath
Amaurobius similis	SV253		03043004.jpg	Amaurobiidae/Amaurobius/A_similis/ff	...
Amaurobius similis	SR276		03043004.jpg	Amaurobiidae/Amaurobius/A_similis/ff	...
Amaurobius fenestralis	SR130	drawing of heart, intestine	Millot_1938c_fig_45.jpg	Amaurobiidae/Amaurobius/A_fenestralis/ff	...
Amaurobius fenestralis	SR47	drawing of heart, intestine	Millot_1938c_fig_45.jpg	Amaurobiidae/Amaurobius/A_fenestralis/ff	...

FIGURE 3. The phylogenetic matrix table (matrix cells) and the four tables used by the SILK package for Mesquite to query for images to display in cells. Images can be accessed by standard view identifiers (SV numbers) or by anatomical terms (SR numbers) in case they are non-standard images. Fields marked with * are not used for queries or display, only for debugging purposes. Text between parentheses is added here for explanatory purposes only.

FIGURE 4. Images in cells, as displayed by the SILK package for Mesquite. Color density reflects number of images. The left image pane shows the image associated with the current cell, the right pane shows exemplar or "type" images used to illustrate the definition of character states.

standard views. One may wish to retrieve all images conceivably displaying the homologue, regardless of magnification, device, or technique. Higher level ontological relations make it relatively easy to expand or contract the scope of the query or to toggle between queries, and a recursive query can search parent anatomical regions in the ontology until images are found. Relationships defined ontologically as serial or modular homologues would likewise enable retrieval of images documenting all setae whether they are hairs, scales, trichobothria or macrosetae, all tarsal claws, or all spigots on different spinnerets. Ontologies can also represent behaviors (e.g., http://www.ethodata.org/; Midford, 2004) in which one or more homologues may be involved, such as stridulation or silk spinning, and therefore retrieve all images that pertain to such behaviors.

DISTRIBUTION OF DATA, THE SEMANTIC WEB AND INFORMAL TAGGING

Distributed data maintenance enormously accelerates the accumulation of knowledge, because different pieces of information can be updated over time without depending on a given research group. This requires that the object identifiers are globally unique and durable, with the associated metadata easily accessible, as foreseen for the Semantic Web project (http://www.w3.org/2001/sw/; see Page, 2006). Until such identifiers and metadata services are in place, our approach will rely on relational databases. Our system has a number of similarities to Semantic Web projects, especially our use of ontology-based inferences to locate stored images. These similarities follow from a shared interest in correctness, both to avoid naming ambiguities and to assure proper inference. If we were to add web-based image searches to this system, we could serve RDF-format annotations along with the images. Those images would then be free-form searchable by the community without need to query our databases.

We have identified points in our workflow where the input of metadata is both economic and reliable, because the participant is focused in the problem and has the relevant materials at hand. The most obvious are the time of creation of objects (production of an image, insertion of a term in the ontology, or of a new character), but not the only ones. A transverse pass is a good moment to tune up associations between standard views and the character, and review the specifications of standard views itself; the scoring of the data set is the best moment to mark observations that challenge the definition of a character. In the long term we expect that further metadata will be continuously added or reviewed by users, in a diversity of contexts, including the submission and annotation of legacy images. These additions may introduce issues of scaling in our system. Collaborative tagging is a promising solution (Golder and Huberman, 2005), and we expect that a well curated and documented ontology of homologues will provide the participants with the tools for consistent and accurate tagging in the vast majority of cases. Free-form tags may serve the fraction of terms not supported by the ontology and would be a valuable source for updates and additions to the ontology.

CONCLUSIONS

Although the obstacles to progress and synthesis in comparative morphology that we identify are by no means new, the techniques and conceptual framework proposed here offer at least partial remedies based on new technologies and approaches. We see the need for robust ontologies that strictly reflect known and hypothesized homology relationships as fundamental to the interaction of collections of images, characters, taxa, and specimens, and therefore to the efficient workflow of large, distributed, multicollaborator long-term phylogenetic projects. Such ontologies must be sufficiently rigorous to support machine-processing of large amounts of comparative data and images.

Perhaps the most significant benefit is more efficient and intelligent exploratory tools. As difficult as it is to produce high-quality comparative morphological data, it is still more difficult to organize, store, retrieve, filter, and synthesize it, and the problem will only worsen. Large projects with multiple collaborators require flexible subdivision into more or less stand-alone components than can proceed in parallel and independently. Data management must gracefully facilitate late-stage data production and continual updates of character definitions, cell scores, and taxonomic changes. The project as a whole should prefigure the distributed network of global repositories of biological data already under construction. Metadata should permanently link observations to specimens, as already implemented in initiatives such as MorphBank and MorphoBank. Finally, insofar as possible, data collection and categorization should be as standardized as possible to facilitate large-scale distributed machine-processing now and in the future.

REFERENCES

Agosti, D., and N. F. Johnson. 2002. Taxonomists need better access to published data. Nature 417:222.

Bisby, F. A., J. Shimura, M. Ruggiero, J. Edwards, and C. Haeuser. 2002. Taxonomy, at the click of a mouse. Nature 418:367.

Blanco, W., C. Gaitros, D. Gaitros, N. Jammigumpula, K. Maneva-Jakimoska, D. Paul, F. Ronquist, K. Seltmann, and S. Winner. 2006. MorphBank v.2.2 user manual, v. 9 May 2006. http://morphbank.net/.

Brusca, R. C., and G. J. Brusca. 2002. Invertebrates. Sinauer Associates, Sunderland, Massachusetts.

Coddington, J. A. 1990. Ontogeny and homology in the male palpus of orb weaving spiders and their relatives, with comments on phylogeny (Araneoclada: Araneoidea, Deinopoidea). Smithson. Contrib. Zool. 496:1–52.

Day-Richter, J. 2001-2006. OBO-Edit. An open source ontology editor. http://sourceforge.net/.

Dettai A., N. Bailly, R. Vignes-Lebbe, and G. Lecointre. 2004. Metacanthomorpha: Essay on a phylogeny-oriented database for morphology—The acanthomorph (Teleostei) example. Syst. Biol. 53:822–834.

Driskell, A. C., J. G. Burleigh, M. M. McMahon, B. C. O'Meara, and M. J. Sanderson. 2004. Prospects for building the tree of life from large sequence databases. Science 306:1172–1174.

Gaston, K. J., and R. M. May. 1992. Taxonomy of taxonomists. Nature 356:281–282.

Gewin, V. 2002. All living things, online. Nature 418:362–363.

Godfray, H. C. J. 2002a. Challenges for taxonomy. Nature 417:17–19.

Godfray, H. C. J. 2002b. Towards taxonomy's "glorious revolution." Nature 420:461.

Godfray, H. C. J., and S. Knapp. 2004. Taxonomy for the 21st century: Introduction. Phil. Trans. R. Soc. Lond. B 359:559–569.

Golder, S., and B. A. Huberman. 2005. The structure of collaborative tagging systems. http://arxiv.org/pdf/cs.DL/0508082.

Goloboff, P. A. 1999. Analyzing large data sets in reasonable times: Solutions for composite optima. Cladistics 15:415–428.

Griswold, C. E. 1993. Investigations into the phylogeny of the Lycosoid spiders and their kin (Arachnida, Araneae, Lycosoidea). Smithson. Contrib. Zool. 539:1–39.

Griswold, C. E., J. Coddington, G. Hormiga, and N. Scharff. 1998. Phylogeny of the orb-web building spiders (Araneae, Orbiculariae: Deinopoidea, Araneoidea). Zool. J. Linn. Soc. 122:1–99.

Griswold, C. E., M. J. Ramírez, J. Coddington, and N. Platnick. 2005. Atlas of Phylogenetic Data for Entelegyne spiders (Araneae: Araneomorphae: Entelegynae) with comments on their Phylogeny. Proc. Calif. Acad. Sci. 4th Ser. 56 II:1–324.

Guarino, N., and C. A. Welty. 2004. An overview of Ontoclean. Pages 151–172 in The handbook on ontologies (S. Staab and R. Studer, eds). Springer-Verlag, Berlin.

Hormiga, G. 1994a. A revision and cladistic analysis of the spider family Pimoidae (Araneae: Araneoidea). Smithson. Contrib. Zool. 549:1–105.

Hormiga, G. 1994b. Cladistics and the comparative morphology of linyphiid spiders and their relatives (Araneae, Araneoidea, Linyphiidae). Zool. J. Linn. Soc. 111:1–71.

Janies, D. A., and W. C. Wheeler. 2001. Efficiency of parallel direct optimization. Pages S71–S82 in One day symposium in numerical cladistics (G. Giribet, W. C. Wheeler, and D. A. Janies, eds.). Cladistics 17:S71–S82.

Jenner, R. A. 2001. Bilaterian phylogeny and uncritical recycling of morphological data sets. Syst. Biol. 50:730–742.

Klaus, A. V., V. L. Kulasekera, and V. Schawarock. 2003. Three-dimensional visualization of insect morphology using confocal laser scanning microscopy. J. Microsc. 212:107–121.

Maddison, D. R., and W. P. Maddison. 2000. MacClade 4: Analysis of phylogeny and character evolution. Sinauer Associates, Sunderland, Massachusetts.

Maddison, W. P., and D. R. Maddison. 2006. Mesquite: A modular system for evolutionary analysis, version 1.12. Available online at http://mesquiteproject.org/.

Maddison, W. P., and M. J. Ramírez. 2006. Simple Image LinKing (SILK): A Mesquite package for associating images with character matrices. Beta test version available online at http://mesquiteproject.org/SILK/.

Midford, P. E. 2004. Ontologies for behavior. Bioinformatics 20:3700–3701.

Nixon, K. C. 1999a. The parsimony ratchet, a new method for rapid parsimony analysis. Cladistics 15:407–414.

Nixon, K. C. 1999b. Winclada (v. 1.00.04). Published by the author, Ithaca, New York. Available at http://www.cladistics.com/about_winc.htm.

Nixon, K. C., J. Carpenter, and S. Borgardt. 2001, Beyond NEXUS: Universal cladistic data objects. Cladistics 17:S53–S59.

Page, R. D. M. 2001. Nexus Data Editor for Windows (NDE), version 0.5.0. Program and documentation. Available online at: http://taxonomy.zoology.gla.ac.uk/rod/NDE/nde.html. Published by the author, Glasgow, UK.

Page, R. D. M. 2004a. Phyloinformatics: Towards a phylogenetic database. Pages 219–241 in Data mining in bioinformatics (J. T. L. Wang, M. J. Zaki, H. T. T. Toivonen, and D. Shasha, eds.). Springer-Verlag, Berlin.

Page, R. D. M. 2004b. Taxonomy, supertrees, and the Tree of Life. Pages 247–265 in Phylogenetic supertrees: Combining information to reveal the tree of life (O. Bininda-Emonds, ed.). Kluwer, Amsterdam, The Netherlands.

Page, R. D. M. 2005. A taxonomic search engine: Federating taxonomic databases using web services. BMC Bioinformat. 6:48.

Page, R. D. M. 2006. Taxonomic names, metadata, and the semantic web. Biodivers. Informat. 3:1–15.

Platnick, N. I. 2006. The world spider catalog, version 7.0. The American Museum of Natural History. http://research.amnh.org/entomology/spiders/catalog/index.html.

Platnick, N. I., J. A. Coddington, R. R. Forster, and C. E. Griswold. 1991. Spinneret morphology and the phylogeny of haplogyne spiders (Araneae, Araneomorphae). Am. Mus. Novit. 3016:1–73.

Prendini, L. 2000. Phylogeny and classification of the Superfamily Scorpionoidea Latreille 1802 (Chelicerata, Scorpiones): An exemplar approach. Cladistics 16:1–78.

Prendini, L. 2001. Species or supraspecific taxa as terminals in cladistic analysis? Groundplans versus exemplars revisited. Syst. Biol. 50:290–300.

Proszynski, J. 2003–2006. Salticidae (Araneae) of the World, version March 1st, 2006. Museum and Institute of Zoology, Polish Academy of Sciences, online at http://salticidae.org/salticid/diagnost/title-pg.htm.

Ramírez, M. J. 2000. Respiratory system morphology and the phylogeny of haplogyne spiders (Araneae, Araneomorphae). J. Arachnol. 28:149–157.

Raven, R. J., and K. Stumkat. 2005. Revisions of Australian ground-hunting spiders: II: Zoropsidae (Lycosoidea: Araneae). Mem. Queensl. Mus. 50:347–423.

Rodman, J. E., and J. H. Cody. 2003. The taxonomic impediment overcome: NSF's partnerships for enhancing expertise in taxonomy (PEET) as a model. Syst. Biol., 52:428–435.

Ronquist F., and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574.

Rosse, C., and J. L. V. Mejino. 2003. A reference ontology for bioinformatics: The foundational model of anatomy. J. Biomed. Informat. 36:478–500.

Sattler, R., 1996. Classical morphology and continuum morphology: Opposition and continuum. Ann. Bot. 78:577–581.

Scharff, N., and J. A. Coddington. 1997. A phylogenetic analysis of the orb-weaving spider family Araneidae (Arachnida, Araneae). Zool. J. Linn. Soc. 120:355–434.

Schütt, K. 2003. Phylogeny of Symphytognathidae s.l. (Araneae, Araneoidea). Zool. Scripta 32:129–151.

Shadbolt, N., T. Berners-Lee, and W. Hall. 2006. The Semantic Web revisited. IEEE Intelligent Systems 21:96–101.

Silva Dávila, D. 2003. Higher-level relationships of the spider family Ctenidae (Araneae: Ctenoidea). Bull. Am. Mus. Natl. Hist. 274:1–86.

Smith, B. 2005. The logic of biological classification and the foundations of biomedical ontology. Pages 505–520, in Petr Hájek, Luis Valdés-Villanueva and Dag Westerståhl (ed.), Logic, methodology and philosophy of science. Proceedings of the 12th international conference, London, King's College Publications.

Smith, B. 2004b. The logic of biological classification and the foundations of biomedical ontology. Pages in Invited Papers from the 10th International Conference in Logic Methodology and Philosophy of Science, Oviedo, Spain, 2003 (D. Westerståhl, ed.). Elsevier-North-Holland, Amsterdam, The Netherlands.

Smith, B., W. Ceusters, B. Klugges, J. Köhler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, A. L. Rector, and C. Rosse. 2005. Relations in biomedical ontologies. Genome Biol. 6:R46.

Soltis, D. E., P. S. Soltis, P. K. Endress, and M. W. Chase. 2005. Phylogeny and evolution of angiosperms. Sinauer Associates, Sunderland, Massachusetts.

Stamatakis, A., T. Ludwig, and H. Meier. 2005. RAxML-III: A fast program for maximum likelihood-based inference of large phylogenetic trees. Bioinformatics 21:456–463.

Systematics Agenda 2000. 1994. Systematics Agenda 2000: Charting the biosphere. Technical report. Systematics Agenda 2000, Society of Systematic Biologists, Willi Hennig Society, and Association of Systematics Collections, New York.

Thacker, P. D. 2003. Morphology: The shape of things to come. BioScience, 53:544–549.

Trelease, R. B. 2006. Anatomical reasoning in the informatics age: Principles, ontologies, and agendas. Anat. Rec. (New Anat.) 289b:72–84.

Westphal, M. 2006. Imatch 3.5.0.22. Available at http://www.photools.com/index.php.

Wheeler, Q. D. 2003. Transforming taxonomy. Systematist 22:3–5.

Wheeler, Q. D. 2004. Taxonomic triage and the poverty of phylogeny. Phil. Trans. R. Soc. Lond. B 359:571–583.

Wilson, E. O. 2003. The encyclopedia of life. Trends Ecol. Evol. 18:77–80.

Wilson, E. O. 2004. Taxonomy as a fundamental discipline. Phil. Trans. R. Soc. Lond. B 359:739.

Wirkner, C. S., and S. Richter. 2004. Improvement of microanatomical research by combining corrosion casts with MicroCT and 3D reconstruction, exemplified in the circulatory organs of the woodlouse. Microsc. Res. Tech. 64:250–254.
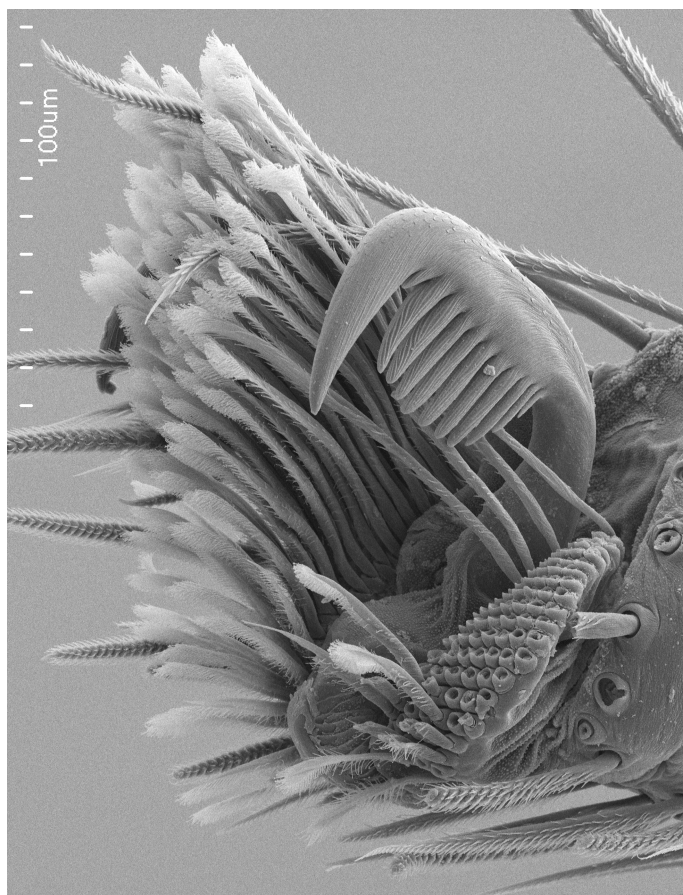
Tarsal claws of the salticid spider *Cocalodes longicornis* from New Guinea (left leg I, preparation MJR-00444, AMNH). When available electronically, high resolution scanning electron microscope images can be used to explore characters beyond the original purpose of the image. When zooming in, this example shows fine details such as barbs on setae, the shape of setal sockets, and the sculpture of the cuticle.