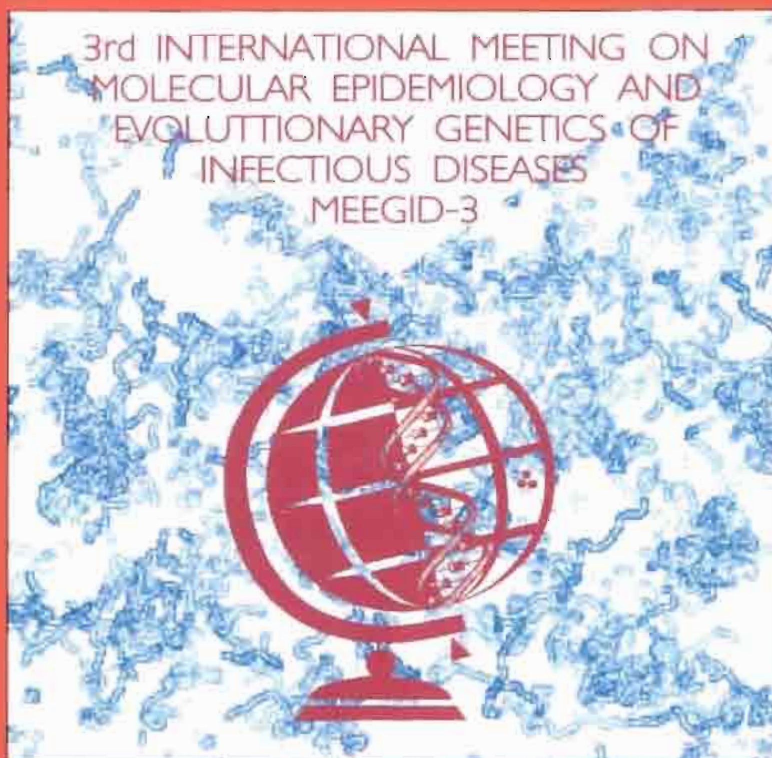
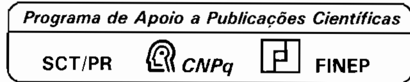
Vol. 93(5) Sep./Oct. 1998

MEMÓRIAS DO INSTITUTO OSWALDO CRUZ

An International Journal of
Biological and Biomedical Research



MEMÓRIAS DO INSTITUTO OSWALDO CRUZ



©1998 Fundação Oswaldo Cruz - Fiocruz
Ministério da Saúde

Mem. Inst. Oswaldo Cruz is covered by

Biological Abstracts
Current Contents - Life Sciences
Entomology Abstracts
Excerpta Medica
Helminthological Abstracts
Index Medicus
Index Medicus Latino-Americano
Lit. Latino-Americana - LILACS
Microbiology Abstracts
Protozoological Abstracts
Review of Applied Entomology
Science Citation Index
Tropical Diseases Bulletin
Zoological Abstract

Articles from this journal are available through

- the authors
- the Genuine Article®, 3501 Market Street, Philadelphia, PA 19104, USA
- the Biosis Document Express, 41575 Joy Road, Canton, MI 48187-2062, Fax: 313-459.8990
  E-mail: biosisdoc@advinfoc.com
- SourceOne Uncover, Fax: 303-758.5946 or internet: uncover@carl.org

On line access

The address of the home page of the Memórias is:
http://www.pobox.com/~memorias
where the full text of articles can be accessed.

Editorial Office

Administrative Editors	Luciane CB Soares
Consulant	Marly de C Willcox
Administrative Office	Wanilda Blandy Correa
	Geiza Helena R Neves
Desktop Publishing	Paulo Márcio Moreira
On line version:	Leonardo M Rocha

Subscription information

The subscription price for one volume of six issues including postage is:

- Institutional - Brazil R$ 70,00, South America US$ 80.00, rest of world US$ 120.00.
- Personal - Brazil R$ 35,00, South America US$ 45.00, rest of the world US$ 60.00.

It is possible for Institutions which sponsor scientific journals to exchange issues with the Memórias in lieu of payment.

All manuscripts and correspondence concerning editorial matters, subscriptions, and changes of address should be sent to

Memórias do Instituto Oswaldo Cruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil
Phone: +55-021-598.4335 Fax: +55-21-280.5048 E.mail: memorias@pobox.com

MEMÓRIAS DO INSTITUTO OSWALDO CRUZ

ISSN-0074-0276

President of FIOCRUZ: Eloi de Souza Garcia
Director of the Instituto Oswaldo Cruz: José Rodrigues Coura
Editor: Hooman Momen

Editorial Board (Rio de Janeiro)

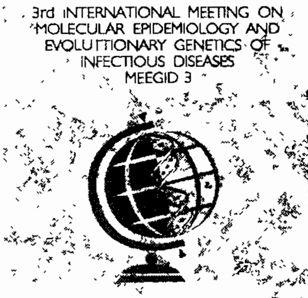
Ana Carolina Vicente
Carlos André Salles
Carlos M Morel
Claude Pirmez
Claudio J Struchiner
Claudio T Daniel Ribeiro
Delir Corrêa Gomes
Eloi S Garcia
Gabriel Grimaldi Jr
Herman G Schatzmayr
José Rodrigues Coura
Maria de Fatima Cruz
Maria Nazareth L Meirelles
Marli Maria Lima
Renato S Balão Cordeiro
Ricardo L de Oliveira
Yara M Traub-Cseko

International Advisory Board

Alberto CC Frasch (Buenos Aires)
Alain de Chambrier (Geneva)
Allan A Yousten (Blacksburg)
Allen W Cheever (Bethesda)
Ana Flisser (Mexico)
Anthony J Shelley (London)
Antoniana U Krettli (Belo Horizonte)
Bianca Zingales (São Paulo)
Brian WJ Mahy (Atlanta)
Daniel Colley (Atlanta)
Daniel Camus (Villeneuve d'Asc)
David H Molyneux (Liverpool)
Diane McMahon-Pratt (New Haven)
Eric Milstrey (Fort Polk)
Franklin A Neva (Bethesda)
Gerusa Dreyer (Recife)
Humberto Guerra (Lima)
JD Thomas (Brighton)
Jeffrey Shaw (Brasília)
John R David (Boston)
Jorge Kalil (São Paulo)
José Luis Ramírez Ochoa (Caracas)
KP Chang (Chicago)
Luis Travassos (São Paulo)
M Goreti Rosa-Freitas Sibajev (Boa Vista)
Philippe Desjeux (Geneva)
R Killick-Kendrick (Ascot)
Ralph Harbach (London)
Ralph Lainson (Belém)
Ruth S Nussenzweig (New York)
Rodrigo Zeledón (San José)
Socrates Herrera (Cali)
Stephen M Beverley (St Louis)
Terry Klein (Washington)
Thomas C Cheng (Charleston)
Toby Barrett (Manaus)
Victor Nussenzweig (New York)
Wallace Peters (St. Albans)
Walter A Boeger (Curitiba)
Zigman Brener (Belo Horizonte)

Publication date

Vol. 93 (5) .... 10 September 1998



3rd International Meeting on Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases.

© 1998 Fundação Oswaldo Cruz - Fiocruz. All rights reserved. ISSN 0074-0276/1998/ $4.00

This journal and the individual contributions contained in it are protected by the copyright of Fiocruz, and the following terms and conditions apply to their use:

Photocopying
Single photocopies of single articles may be made for personal use as allowed by national copyright laws. Permission of the publisher and payment of a fee is required for all other photocopying, including multiple or systematic copying, copying for advertising or promotional purposes, resale, and all forms of document delivery. Special permissions are available for educational institutions that wish to make photocopies for non-profit educational classroom use. In the USA, users may clear permissions and make payment throught the Copyright Clearance Center Inc. (CCC account number 1000011958), 222 Rosewood Drive, Danvers MA 01923, USA. Tel. (508) 750-8400; Fax: (508) 750-4744. In other countries where a local copyright clearance centre exists, please contact it for information on required permission and payments.

Derivative Works
Subscribers may reproduce tables of contents or prepare lists of articles including abstracts for internal circulation within their institutions. Permission is required for resale or distribution outside the institution. Permission is required for all other derivative works, including compilations and translations. Permission is required to store electronically any material contained in this journal, including any article or part of an article. Except as outlined above, no part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, eletronic, mechanical, photocopying, recording or otherwise, without prior written permission.

FICHA CATALOGRÁFICA

Memórias do Instituto Oswaldo Cruz. v. 1, 1909- Rio de Janeiro: Fiocruz: Secretaria das Memórias, 1909 - v.: il.; 27cm.

Irregular: 1909-1976.
Interrompida: 1976-1980.
Trimestral. 1981-1994.
Bimestral a partir do Vol. 90, 1995.
ISSN 0074-0276.

1. Medicina experimental - Periódicos - Brasil. 2. Pesquisa científica - Periódicos - Brasil.
I. Fundação Oswaldo Cruz. II. Instituto Oswaldo Cruz. Secretaria das Memórias.

CDD - 001.4305
- 619.05

CDU - 61.001.6(81)(05)
- 001.891/.2(81)(05)

3rd International Meeting on Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases

7-10 June 1998 - Rio de Janeiro, RJ

FOREWORD

The Third International Workshop on Molecular Epidemiology and Evolutionary Genetics of Infectious Diseases was held at the Hotel Glória in Rio de Janeiro, Brazil, from June 7 to 10, 1998. The title of this third meeting was broadened to cover infectious diseases so as to include both vector and host aspects as well as pathogenic micro-organisms.

The 11 plenary lectures and 14 round-tables presented during this workshop covered a wide variety of diseases from a number of different perspectives. The abstracts received from over 20 countries and six continents attested to the popularity and widespread appeal of these meetings. Brazil was an appropriate setting for this meeting as most of the infectious diseases discussed during this workshop are either emerging, re-emerging or endemic in this country. These international meetings started from an idea shared between Michel Tibayrenc and Altaf Lal. The first meeting was held in June 1996 in Atlanta GA, USA and the second in Montpellier in May 1997. These two meetings were co-sponsored by ORSTOM (the National French Agency for scientific research in developing countries), CNRS (the National French Agency for basic research) and the Centers for Disease Control and Prevention (CDC). For this third meeting the Oswaldo Cruz Institute of the Oswaldo Cruz Foundation joined the original sponsors. The Oswaldo Cruz Institute was founded in Rio de Janeiro in the beginning of this century and has a distinguished record of achievements in the field of research and control of infectious diseases. Since these meetings were founded the importance of the molecular epidemiological and evolutionary genetic approach to infectious diseases has been increasingly demonstrated in the identification and control of many outbreaks. Several practical examples of the use of this approach were given in the talks during the meeting. The full program and abstracts of all the presentations (plenary lectures, round-tables and posters) are available at the web-site for the event http://www.dbbm.fiocruz.br/www-mem/meeting. In addition the speakers of the oral presentations were invited to submit manuscripts to be considered for publication in the *Memórias*. In order for the manuscripts to be published shortly after the meeting a deadline was imposed for the submission of the manuscripts. Due to the short time available many speakers were unable to make submissions, however those who sent manuscripts and which were approved for publication are included in this issue of the journal.

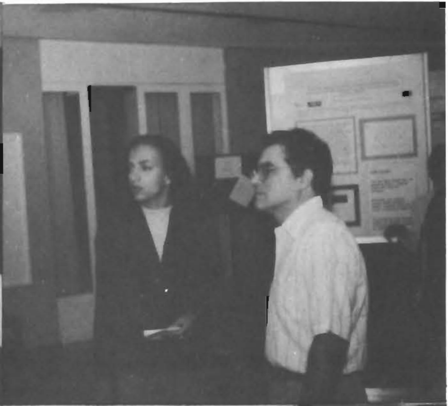
We would like to thank the following organizations for their financial support of this meeting: CNPq (The Brazilian National Research Council), FAPERJ (The State of Rio de Janeiro Research Council), CAPES (The Brazilian Agency for post-graduate studies), FNS (The Brazilian National Health Foundation), INTERACTIVA Biotechnologie Gmbh and Sigma Chemical Co. (Brazil). We would also like to acknowledge the support of the Brazilian societies of Mycology, Virology and Microbiology.

From the many comments received both during and after the workshop it can be concluded that the meeting was very successful, both in terms of the high quality of the presentations and in the opportunities provided by the intervals and social program for contacts and interactions among the participants. The National press also took great interest in the workshop and articles appeared in newspapers and magazines, before, during and after the meeting as well as material, on television news and radio.

The success of this meeting bodes well for the next workshop which is planned for Dakar, Senegal in June 1999. Further information about this meeting can be obtained from Dr Michel Tibayrenc (fax: +33-4-67416299) or from the organizers below.

The organizers

Hooman Momen
(hmomen@gene.dbbm.fiocruz.br)
Altaf A Lal
(aall@cdc.gov)
Michel Tibayrenc
(Michel.Tibayrenc@cepm.mpl.orstom.fr)



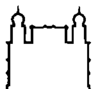
CDC

DR. ELDI GARCIA

3rd INTERNATIONAL MEETING ON MOLECULAR EPIDEMIOLOGY AND EVOLUTIONARY GENETICS OF INFECTIOUS DISEASES

MEEGID-3

HOTEL GLÓRIA
RIO DE JANEIRO, JUNE 7-10, 1998



Ministério da Saúde
Fundação Oswaldo Cruz
Instituto Oswaldo Cruz

ORSTOM

CDC
Centers for Disease Control and Prevention

CNRS
CENTRE NATIONAL DE LA RECHERCHE SCIENTIFIQUE

Evolutionary Control of Infectious Disease: Prospects for Vectorborne and Waterborne Pathogens

Paul W Ewald/+, Jeremy B Sussman, Matthew T Distler, Camila Libel, Wahid P Chammas, Victor J Dirita*, Carlos André Salles, Ana Carolina Vicente***, Ingrid Heitmann****, Felipe Cabello*****

Department of Biology, Amherst College, Amherst, MA 01002-5000, USA *Laboratory of Animal Medicine, University Michigan School of Medicine, Ann Arbor, MI 48109, USA **Departamento de Bioquímica e Biologia Molecular ***Departamento de Genética, Instituto Oswaldo Cruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil ****Sub Departamento de Microbiologia Clinica, Instituto de Salud Publica, Maraton 1000, Nunoa Santiago, Chile *****Department of Immunology, New York Medical College, Vahalla, NY, USA

Evolutionary theory may contribute to practical solutions for control of disease by identifying interventions that may cause pathogens to evolve to reduced virulence. Theory predicts, for example, that pathogens transmitted by water or arthropod vectors should evolve to relatively high levels of virulence because such pathogens can gain the evolutionary benefits of relatively high levels of host exploitation while paying little price from host illness. The entrance of Vibrio cholerae *into South America in 1991 has generated a natural experiment that allows testing of this idea by determining whether geographic and temporal variations in toxigenicity correspond to variation in the potential for waterborne transmission. Preliminary studies show such correspondences: toxigenicity is negatively associated with access to uncontaminated water in Brazil; and in Chile, where the potential for waterborne transmission is particularly low, toxigenicity of strains declined between 1991 and 1998. In theory vector-proofing of houses should be similarly associated with benignity of vectorborne pathogens, such as the agents of dengue, malaria, and Chagas' disease. These preliminary studies draw attention to the need for definitive prospective experiments to determine whether interventions such as provisioning of uncontaminated water and vector-proofing of houses cause evolutionary reductions in virulence.*

Key words: infectious diseases - control - pathogens - waterborne transmission

AN EVOLUTIONARY APPROACH TO VIRULENCE

The ongoing synthesis of epidemiology, molecular biology, and evolutionary biology promises to improve our understanding of the temporal and geographic variation in pathogens and the diseases they cause. From a practical viewpoint this improved understanding may prove useful in identifying new possibilities for the control and prevention of infectious disease. One aspect of the infectious process that seems particularly amenable to this control is virulence, which is defined here as the level of harm to the host. Although virulence depends on the interplay between pathogen and host characteristics, it is useful to consider the inherent virulence of a pathogen as the pathogen's contribution to this harmfulness. In practice this contribution is not separable from the host in which the harmfulness is assessed, yet conceptually reference to the inherent virulence of pathogens in the context of the spectrum of infectious agents. The smallpox viruses are inherently more harmful than rhinoviruses even though some of the mildest smallpox virus infections may be no more severe than the most severe rhinovirus infections.

Evolutionary considerations emphasize that the inherent virulence of pathogens should depend on a tradeoff between fitness benefits and fitness costs that are associated with particular levels of virulence. The fitness benefits are accrued through increased replication of the genetic instructions for the characteristic. Costs are typically accrued through reductions in the transmission of the genetic instructions, for example, due to negative effects of host illness on pathogen transmission. Evolutionary theory generally does not propose that virulence per se is beneficial. Rather, the logic

This study was supported by two grants from Leonard X Bosack and Bette M Kruger Charitable Foundation (PWE), an Amherst College Faculty Research Award (PWE), a Hughes student fellowship (CL), and the Webster Fund of the Biology Department at Amherst College.
+Corresponding author. Fax: +413-542.7955
Received 15 June 1998
Accepted 30 July 1998

assumes that disease organisms may benefit by exploiting their hosts. Such exploitation allows a disease organism to secure resources that it can use to reproduce, and thereby contribute more copies of the instructions for that exploitation into future generations. These fitness benefits of exploitation are weighed against the costs. The illness caused by intense levels of exploitation may make the host immobile, host mobility may be necessary for transmission to new hosts (as is the case, for example, with the common cold virus). In this case pathogen variants that exploit hosts so intensely that they cause host immobility may get more resources in the short run, but lose in the slightly longer run because of reductions in transmission. Pathogens that do not rely on host mobility for transmission pay a relatively low price if their exploitation immobilizes the host. According to the tradeoff reasoning presented above, pathogens in such categories should be particularly virulent. One of these categories involves waterborne transmission.

WATERBORNE TRANSMISSION

Waterborne transmission allows diarrheal pathogens to be transported from immobilized infected hosts to uninfected hosts. Where water supplies are not protected, a person with incapacitating diarrheal illness will release the diarrheal pathogens into clothes, bed sheets, or containers for collecting excreta. These items then tend to be removed by attendants and washed in bodies of water such as canals or rivers, which may be used as sources of drinking water or may flow into supplies of drinking water. Either way, the cycle is completed when susceptible individuals drink the contaminated water. In this situation, hightly exploitative (and hence highly virulent) pathogen variants should be favored by natural selection because the benefits of intense exploitation are great and the costs of exploitation are small. The benefits are great because large numbers of susceptibles can be infected by the increased numbers of propagules in the water. The costs are low because the incapacitating illness associated with this propagule production should have relatively little negative effect on the waterborne transmission of the propagules-rather than relying on the mobility of the infected individuals to enact transmission, the pathogens are using the mobility of the attendants and the water.

This hypothesized effect of waterborne transmission has been tested by determining whether the lethality of bacterial agents of human diarrhea is positively correlated with the degree to which they are waterborne (Ewald 1991). Fig. 1 shows that this correlation exists. Variation in the virulence of human diarrheal diseases can thus be explained in an evolutionary sense by variation in the degree to which different diarrheal pathogens are waterborne. This association offers some insight into the variation in virulence that occurs among diarrheal bacteria, but perhaps more importantly it suggests a new means for lessening the damage associated with diarrheal diseases. By reducing the potential for waterborne transmission we may be able to force diarrheal pathogens to evolve reduced virulence.

Whether this possibility is feasible depends on the validity of applying the trend apparent across the broad spectrum of diarrheal pathogens depicted in Fig. 1 to particular pathogens. Would a particular kind of pathogen evolve reduced virulence in response to a reduced potential for waterborne transmission? If so, what time period would be

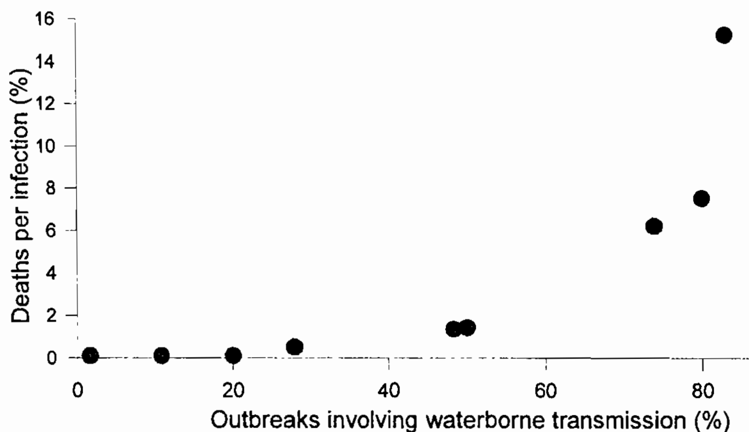


Fig. 1: waterborne transmission and mortality of diarrheal bacteria of humans. Pathogens ordered from most to least waterborne are classical *Vibrio cholerae*, *Shigella dysenteriae* type 1, *Salmonella typhi*, el tor *V. cholerae*, *Shigella flexneri*, *Shigella sonnei*, enterotoxigenic *Escherichia coli*, *Campylobacter jejuni*, and nontyphoid *Salmonella* (for other details see Ewald 1991).

required? If this period were a few years, then efforts to improve access to clean water supplies would have evolutionary effects over an interval that is comparable to the intervals envisioned for conventional, nonevolutionary interventions.

V. cholerae is an excellent study subject for this kind of analysis particularly because its virulence is largely attributable to its toxin production. The inherent virulence of particular strains therefore can be assessed by quantifying levels of toxin production *in vitro*. Toxin production generates an efflux of fluid into the small intestine, which appears to provides two benefits to *V. cholerae*: (1) it flushes out competitors throughout the intestinal tract, allowing *V. cholerae* to pass down and out of the tract intact, and (2) it creates a fluid stool that probably facilitates transmission by contamination of the external environment and dissemination in water supplies. *V. cholerae* can persist in the intestine during this tumult because it can swim and adhere to the intestinal lining. The costs of toxin production include (1) the metabolic costs of producing the toxin and (2) the negative effect of toxin on host mobility and the probability of host death. Death from cholera results primarily from the dehydration which in turn results from the loss of fluid due to the toxin.

The cholera epidemic that has been unfolding in South America during the 1990s offers a natural experiment with which to assess the general theory. The first reported cases occurred in Peru at the beginning of 1991. The interval since then thus allows an assessment of whether any evolutionary effects of waterborne transmission can occur over a time interval comparable to the interval necessary for other categories of interventions such as vaccination or hygienic improvements to reduce the frequency of infection.

Within two years of the first reports of cholera from Peru the descendants of the Peruvian *V. cholerae* had spread from this epicenter throughout most countries of South and Central America (Tauxe et al. 1995). This spread set up a temporal and geographic pattern of infection that may allow detailed testing of the proposed evolutionary association between waterborne transmission and toxigenicity of *V. cholerae*.

We first focused on Brazil because water quality varies throughout Brazil, and the Brazilian Ministry of Health provides summaries of the proportion of the population with access to potable water. Moreover the large size of Brazil offers the potential for *V. cholerae* to evolve in different directions within the country. The first reported case of cholera in Brazil was in April 1991, about 2.5 months after the first reported case in Peru (Tauxe et al. 1995).

Although this analysis is still in progress, the results are consistent with an influence of water quality on virulence. If the mean for each state is used as a separate data point, there is a statistically significant negative association between access to potable water and *V. cholerae* toxigenicity (one-tailed $p<0.05$, Spearman $r_s = -0.62$). These data are, however, preliminary in several respects: (1) additional strains need to be obtained to make the accuracy of each data point more comparable. Some data points are based on multiple isolates others are based on only one isolate; (2) each strain was considered to be an independent data point in the statistical test; however, the degree to which the different data points are independent is unknown. Use of molecular phylogenies should allow the generation of tests that use independent pair-wise comparisons (Harvey & Pagel 1991). This kind of comparison should be feasible eventually, but will probably need to be unusually extensive because nucleotide sequencing and pulsed field gel electrophoretic studies to date have detected almost no variation among the pandemic el tor strains (Salles & Momen 1991, Karaolis et al. 1995); (3) changes in toxigenicity need to be tracked to determine whether harmful strains that enter areas with relatively pure water evolve reduced virulence over time.

Although the data from Brazil suggest that *V. cholerae* has evolved toward a lower level of virulence, they do not indicate how mild it could eventually become in response to cycling in areas with uncontaminated drinking water. To provide such an indication, Fig. 2 also plots the rate of toxin production of strains isolated from Texas and Louisiana in coastal areas of the Gulf of Mexico where *V. cholerae* has been endemic. Zymodeme analysis indicates that these US strains cluster with the el tor strains of *V. cholerae* (rather than strains of the classical biotype) but are only distantly related to these "mainstream" el tor strains (Salles & Momen 1991). They therefore appear to have been present in the US for decades, perhaps being the remnant of a global outbreak of cholera that occurred many decades ago. Their low toxigenicity provides an indication of how low *V. cholerae* toxigenicity could become in an area with uncontaminated of drinking water. Accordingly, although the frequency of seropositivity to *V. cholerae* in local populations in this coastal area of the Gulf of Mexico can be substantial (M.M. Levine, personal communication), cases of cholera there are rare. Only about 50 cases have been reported from this region from 1965 through 1991 (Weber et al. 1994).

We are currently evaluating whether toxigenicity declines over time in regions with a low potential for waterborne transmission after *V. cholerae*

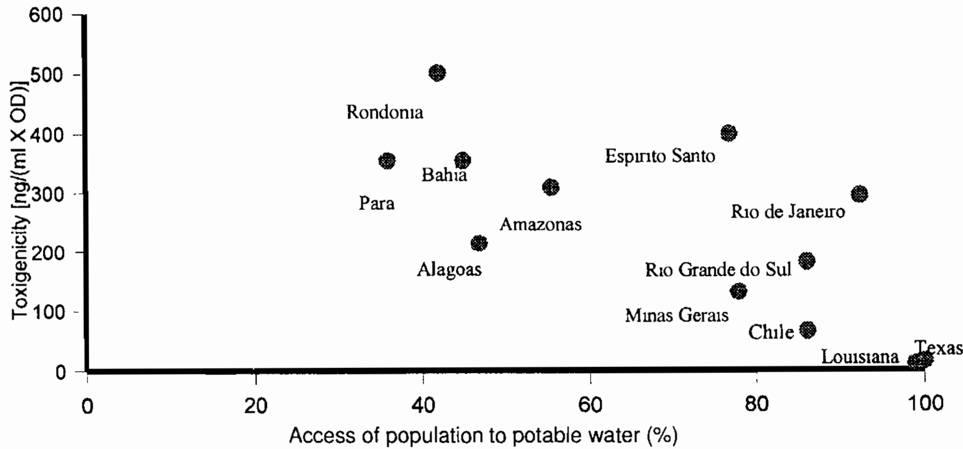


Fig 2. toxigenicities of el tor *Vibrio cholerae* from Brazil, Chile, and the United States. Toxigenicities were assayed using standard ELISA techniques and AKI growth conditions. Names of states are given next to the data point that corresponds to the geometric mean toxigenicity of the strain(s) isolated from the state. Numbers next to each data point refer to the number of different strains tested. About 20 separate measurements of toxigenicity were made for each strain. The geometric mean toxin production was calculated for each strain. When more than one strain was obtained for a state, the geometric mean of the strain means was calculated. Although the *V. cholerae* strains from the United States are only distantly related to those circulating in Brazil, the US strains are presented to provide a sense of how benign strains of *V. cholerae* might become if they are exposed for a long period of time to clean water supplies. The Chilean strains are presented to illustrate how the mean generated from a collection of data associated with a reduction in toxigenicity through time corresponds with the overall geographic trend (see Fig. 2 and text).

enters such areas from regions with a higher potential for waterborne transmission. Our most complete data set in this regard, although still preliminary, comes from Chile. Chile is a particularly important country for evaluation of this hypothesis because it has one of the lowest potentials for waterborne transmission among Latin American countries for which el tor *V. cholerae* infections have become endemic. There is good access to uncontaminated drinking water and a steep elevational gradient that would limit cycles of waterborne transmission. *V. cholerae* entered Chile from Peru at the onset of the pandemic, the first case being reported from Chile about ten weeks after the first case was reported in Peru (Tauxe et al. 1995). Water supplies have been more contaminated in Peru than in Chile since the onset of the South American epidemic.

We have obtained and tested eight Chilean strains isolated from clinical cases over a time span that ranged from 1991 (the first year of the South American epidemic) to the beginning of 1998. The toxigenicity of the tested Chilean strains dropped significantly as a function of time (Fig. 3, one-tailed $p<0.02$, Spearman $r_s=-0.81$).

The geometric mean toxin production of the Chilean strains presented in Fig. 3 is also given in Fig. 2 to allow an assessment of the degree to which their toxigenicity conforms to that of the Brazilian and North American strains (for all data in Fig. 2; Spearman, $p<0.001$, $r_s=-0.75$; $N=11$; the Louisiana and Texas values were treated as a single data point as a conservative measure).

The data from the most recent Chilean isolates are particularly interesting in light of the toxigenicities of the strains isolated in recent years, which are nearly as low as those isolated from the Gulf coast of the US (compare the values for the Chilean isolates of 1998 with those for the Texas and Louisiana). The drop in toxigenicity in Chile corresponds to a very low number of cholera cases. In 1994, for example, when nearly 50,000 cases were reported in Brazil and nearly 25,000 in Peru, the number of reported cases in Chile dropped to one (Tauxe et al. 1995). These figures coupled with the similar differences between the toxigenicities of the US Gulf strains and the most recent Chilean strains further supports the idea that the evolutionary management of virulence is feasible for *V. cholerae* if the potential for waterborne transmission can be sufficiently reduced.

At least four explanations exist for the evolution of reduced toxigenicity of *V. cholerae* in response to reduction in waterborne transmission: (1) the reduction in toxigenicity could result from the increased costs and decreased benefits of toxin production as outlined above; (2) the reduction could result from a variation on this theme, in which the growth of *V. cholerae* in marine environments disfavors toxigenicity, much in the same way the

culturing of parasites outside of hosts causes evolutionary attenuation when genes for virulence no longer provide a fitness benefit to the organism; (3) the decline in Chile could be interpreted as a result of the duration of time that the outbreak had been cycling. Theory suggests that as an outbreak becomes endemic, pathogens might evolve reduced virulence (Lenski & May 1994). To evaluate this hypothesis analogous data are needed from "control" countries invaded by *V. cholerae* at the same time, but for which water quality has remained low. If the reduction in toxigenicity of *V. cholerae* in Chile is attributable at least in part to its low potential for waterborne transmission, this reduction should be stronger than that found in such control countries. We have not yet obtained such a data set, but this comparison is feasible because strains of *V. cholerae* have been isolated in various countries throughout the pandemic; (4) the decrease in waterborne transmission might favor decreased virulence by reducing the genetic heterogeneity of the population of pathogens within a host. Although this hypothesis is probably generally applicable across a broad range of disease organisms, it does not appear to be particularly applicable to *V. cholerae* because its pathogenicity does not involve direct use of host resources, but instead involves the secretion of a product that benefits all of the other *V. cholerae* in the intestinal lumen (see the description of toxin action presented above).

Additional studies are needed to assess these four alternatives. It should be noted however, that from the practical perspective of evolutionary control of disease virulence, the precise mechanism is not so critical as recognition of the association. That is, whatever combination of these explanations is correct, virulence of *V. cholerae* would still evolve toward lower levels in esponse to investments that reduce waterborne transmission.

The comments about phylogenetically paired comparisions mentioned in the context of Fig. 2 also apply to the data in Fig. 3, and neither data set controls for several other variables. Being based on strains that have been isolated and archived, the comparisons do not control, for example, for the source of material. The source of all or virtually all of the strains was clinical material, but the source was often not recorded explicitly in the archived information. Nor was the gathering of strains regimented so as to eliminate gathering biases. Strains isolated at the onset of an outbreak might be disproportionately gathered from severe infections, because severe infections would attract the attention of investigators, who would then develop approaches during the outbreak that would generate samples that were more representative of the existing sample. Although this kind of sampling bias might have contributed to the trend presented in Fig. 3, particularly with regard to the high value for the 1991 isolate, a sampling bias seems inadequate as an explanation of the overall trend, which results from the extremely low levels of toxin production of the strains isolated during the last few years. The toxin production of these strains is one to two orders of magnitude below that of typical el tor strains. Any biases associated with identification of the early cases in the Chilean epidemic should not have created the uniformly low levels of toxin production that were associated with the strains during the latter half of the epidemic; more-

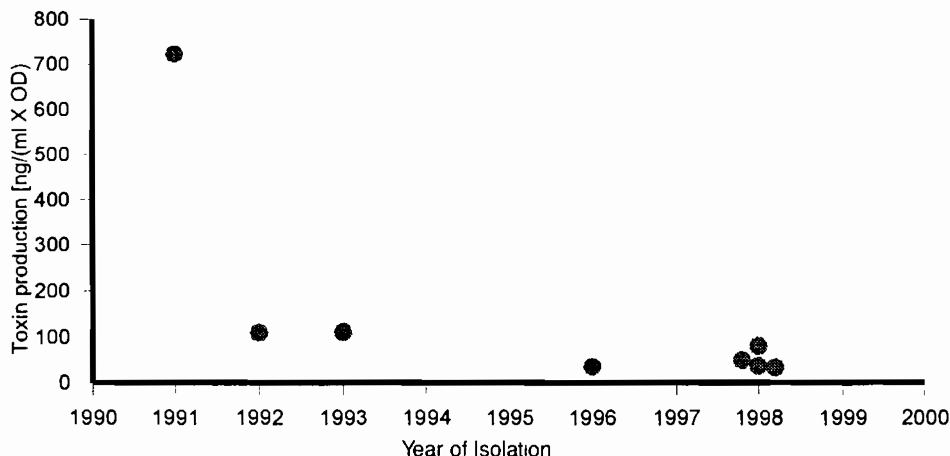


Fig. 3: toxigenicities of el tor *Vibrio cholerae* isolated from Chile from the beginning of the South American outbreak through the beginning of 1998. Each data point corresponds to a different isolate. Figures for access to water supplies are from ministries of health statistics for 1996 (Water access percentages from Louisiana and Texas are artificially separated to allow visualization of data points.). Other details are as described in Fig. 1.

over, the statistical test used is sensitive relative rather than the absolute amounts of toxin production. If the 1991 strain from Chile had, for example, been only one-fifth of its measured value (and substantially less than a "typical" el tor strain), the statistical significance would have remained unchanged.

VECTORBORNE TRANSMISSION

Evolutionary theory identifies vectorborne transmission as a second factor favoring evolution toward relatively high levels of virulence. If a disease organism is transmitted by a biting arthropod vector such as a mosquito or reduviid bug, then it can still be transmitted even if a person is entirely immobilized with illness because such biting arthropods come to feed at immobile people. In fact, experimental studies indicate that mosquitoes are better able to bite a laboratory animal when it is sick with a vectorborne disease such as malaria than when it is healthy, and reduviid bugs (which are vectors for Chagas' disease) typically feed on sleeping individuals. As a consequence, natural selection should favor relatively high levels of host exploitation by vectorborne pathogens, and we should therefore see a particularly high virulence among vectorborne diseases.

The mortality associated with untreated infections is highly variable among both vectorborne and directly transmitted pathogens, but it is greater for vectorborne pathogens than for directly transmitted pathogens (Ewald 1983, 1994). Just as reduction of waterborne transmission should favor evolutionary decreases in virulence, reduction in the potential for arthropodborne transmission from immobilized humans should favor decreases in virulence. This effect can occur through two mechanisms. One mechanism is the direct analog of the argument for waterborne transmission, namely that reduction of transmission from immobilized humans causes a greater reliance on human mobility for transmission. Much as provisioning of uncontaminated drinking water is an intervention that should cause evolutionary reductions in the virulence of diarrheal pathogens, mosquito-proofing of houses is an intervention that should cause an evolutionary reduction in the virulence of vectorborne pathogens such as the agents of malaria or dengue. If a person ill with malaria or dengue stays in bed in a vector-proof house (or hospital), then the transmission of any pathogens in that person will be blocked during that period. To the extent that those variants tend to be inherently more virulent than variants that allow infectious people to be feeling well enough to move around outside of their homes, the composition of the pathogen population will shift toward a greater representation of the milder variants. That is, the pathogen population will have evolved toward mildness. This prediction has not yet been tested, but the information available in the literature, both supports the key steps in logic and suggests that the next stage of large-scale testing is warranted and would be beneficial even if the hypothesis is incorrect.

First, illness tend to be associated with infectiousness. For vectorborne viral diseases, such as dengue, the evidence is straight-forward: viremia occurs during the symptomatic period (e.g., Vaughn et al. 1977). In parasites with more complicated life histories, such as plasmodia, the evidence is more complex because the critical variable is the timing of infectious life history stages (i.e., the gametocytes) is the critical variable. In this case the evidence still supports the idea that much of the transmissibility will be associated with the period of reduced host mobility (e.g., see Ewald 1994).

Geographic variation indicates that parasites have the potential to cause largely mild infections where opportunities for vectorborne transmission are limited. *P. vivax* strains, for example, tend to be more mild in geographic areas associated with low and sporadic mosquito transmission (Ewald 1994). The variation in *P.* vivax's distribution appears to be largely a result of differences in the parasite's tendency to generate dormant resting stages (i.e., "hypnozoites").

P. falciparum infections are often similarly mild where the potential for vectorborne transmission is low, for example, in low transmission areas in the Sudan and Columbia (Elhassan et al 1995, Gonzalez et al. 1997). This tendency also occurs in Mali and more generally along the northern edge of *P. falciparum*'s range in subsaharan Africa (D. S. Peterson, pers. comm.), where the parasite's distribution may be limited by the restricted abundance of mosquitoes. The relative importance of host and parasite characteristics in determining the mildness of *P. falciparum* infections has not been determined in any of these areas, however. If the mildness of such *P. falciparum* infections results at least in part from the mildness of the *P. falciparum* variants, evolution toward reduced virulence would seem particularly feasible. With pre-existing mild strains, detectable evolutionary shifts toward mildness could occur relatively quickly if mosquito-proofing programs were enacted at the edges of *P. falciparum's* distribution. If these programs proved successful the interventions could progress toward the center of the ranges, because the mild strains that would be needed to replace the more severe strains would already be present in the *P. falciparum* gene pool. Although such a progression might facilitate a rapid evolutionary shift to-

ward benignity, it may not be necessary, as variations in pathogen virulence appear to be present even in areas with intense transmission (e.g., Kun et al. 1988).

Influences of exposure to infection on host resistance is a potential confounding variable in any efforts to control malaria through reduction in frequencies of transmission. One hypothesis attracting recent attention proposes that reductions in entomological inoculation rates (EIRs) will have little effect on overall mortality and morbidity in areas with moderate to high bite frequencies, where the benefits of reduced EIRs might be offset by reductions in acquired resistance (Snow & Marsh 1995). With regard to evolutionary effects, this concern is applicable primarily to areas with moderate EIR. In areas with low EIR, mosquito-proofing should lower frequencies of infections to the point of eradication (Watson 1949). In areas with high EIRs, one would expect that mosquito-proofing would cause an evolutionary shift toward benignity with relatively little effect on frequency of infection, and hence with little effect on benefits of acquired immunity. If the evolutionary hypothesis is incorrect, great epidemiological benefits can be expected at least in areas with low EIRs; such nonevolutionary benefits at higher EIRs are uncertain. If the evolutionary hypothesis is correct this benefit at low EIRs will be supplemented with reduced virulence of infections across the spectrum of EIRs.

As is the case with waterborne transmission, vectorproofing of houses can be expected to provide evolutionary reductions in virulence across a spectrum of vectorborne diseases. Different strains of dengue, for example, vary in virulence, with the more virulent strains being more productive in cell culture (Morens et al. 1991). Vector-proofing of houses against dengue's vector, *Aedes aegypti*, should similarly favor the milder less exploitative variants, driving the dengue population to a more benign state. When more than one vectorborne disease is occurring in an area, the overall cost effectiveness may increase in proportion to the number of diseases, because the same intervention should have similar evolutionary effects for each.

The next stage of testing of these ideas will be feasible only if those who control the sources of funds consider the effort worthwhile. The chances of such a positive assessment would be improved if vector-proofing of houses could be shown to have traditional nonevolutionary epidemiological benefits (i.e., reduction in the frequency of infection) in addition to the hypothesized evolutionary epidemiological benefits (i.e., reduction in the harmfulness of the causative organisms). The available evidence indicates that traditional benefits do occur. The effectiveness of mosquito-proof housing against transmission of dengue, for example, is suggested by the resistance to invasion when such housing is generally present. Over the past two decades thousands of cases of dengue fever have occurred on the Mexican side of the US/ Mexico border along the Gulf of Mexico. Dengue has been introduced repeatedly into Texas there but has failed to spread in spite of the ubiquitous presence of *Aedes* vectors. For every reported case acquired on the Texas side of the border there are about 1000 reported cases on the Mexican side (CDC 1996). The pervasiveness of mosquito-proof on the Texas side appears to be responsible for this difference. Similarly, malaria has been introduced on numerous occasions in recent years to areas in the U.S. where it had previously been endemic. Appropriate vectors are abundant, yet little secondary transmission occurs; when it does, it has been self-limited and localized (Wyler 1993, Dawson et al. 1997; for an analogous example involving severe diarrheal disease, see Weissman 1974).

The most thorough experimental test of the effectiveness of mosquito-proof housing on malaria transmission was conducted from 1939 through the 1940s in a large section of northern Alabama, by the Tennessee Valley Authority (TVA), which was overseeing the construction dams in the area (Watson 1949). The TVA was concerned about malaria because the construction of dams in the region had previously contributed to the malaria problem there (Ackerman 1956, Derryberry 1956).

During the 1930s about half of the people in the area tested positive. In 1939, the TVA began a campaign to mosquito-proof all houses in the area and accomplished this goal within seven years. They divided the area into 11 zones and completed the mosquito-proofing of each zone at different times. The results of their study show that mosquito-proofing virtually eradicated malaria from the area, with the decline occurring earlier in those zones in which mosquito-proofing was completed earlier (Fig. 4). No other intervention was enacted prior to the decline (Watson 1949).

These results do not represent a test of the idea that malaria pathogens evolve to lower levels of virulence in response to mosquito-proofing of houses. The results do, however, demonstrate several important points.

First, the results show that *Plasmodium* populations are influenced by mosquito-proofing. If the population as a whole declines so strongly in response to screening, it seems probable that certain variants within the population will be more substantially reduced by screening than others, leading to an evolutionary change in the *Plasmodium* gene pool.

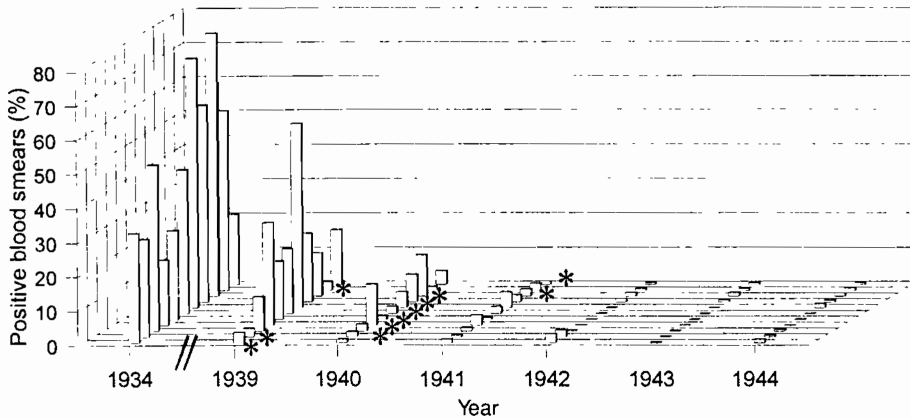


Fig. 4: seropositivity of blood samples for *Plasmodium* presented as a function of year during the mosquito-proofing program carried out in Alabama by the Tennessee Valley Authority. Each row corresponds to one of the 11 geographic zones that comprised the study. The asterisk designates the year in which mosquito-proofing was completed for all houses in the zone. See text for other details (data from Watson 1949).

Second, the results demonstrate nonevolutionary benefits necessary to justify the large-scale evolutionary experiment that would be needed to assess virulence management through mosquito-proofing. To justify the experiments from both ethical and economic perspectives, new areas for experimentation could be selected on the basis of having a slightly more difficult control problem than those for which nonevolutionary success has been demonstrated (e.g., a slightly higher prevalence of infection than occurred in northern Alabama just prior to the mosquito-proofing).

Third, the results show that the experiment is feasible logistically and financially even with the limitations of 1940s technology. The costs of mosquito-proofing (in 1944 dollars) was about $100 per house for the area with the poorest quality of housing; the costs of maintaining the mosquito-proofing was about $12 per house per year (Watson 1949). Modern technology has generated materials that are more effective, more durable, easier to apply and maintain, and more pleasant to live with than those used in the TVA study. Costs should therefore not be as greatly increased as would be indicated by a simple adjustment of the TVA costs for inflation. The actual costs may be influenced up or down depending on the details of a particular area such as the quality of existing houses, the degree to which materials could be generated locally and the costs of local labor.

Finally, the results of the TVA study demonstrate that mosquito-proofing worked even though this geographic area can be stiflingly hot and humid during the malaria season. Skeptics could have argued that people would not stay inside of houses sufficiently under such conditions for the antimalarial effects of mosquito-proof housing to work. Or, skeptics could have argued that people would deliberately destroy screens to increase air-flow through houses, but such vandalism was rare in the Alabama study (Watson 1949).

These ideas should be generally applicable across the spectrum of vectorborne diseases, although the particular details of the application will depend on the details of the vectorborne disease. Chagas disease offers an informative illustration of one variation on the theme. The agent of Chagas disease, *Trypanosoma cruzi*, is transmitted by reduviid bugs that bite sleeping individuals. It is therefore transmitted largely while people are immobilized in their houses. The frequencies of infection should therefore be reduced by vector-proofing of houses.

The details of *T. cruzi* transmission indicate that this intervention could reduce the virulence of *T. cruzi* through two evolutionary processes. The first process is analogous to that proposed above for malaria and dengue. To the extent that transmission does sometimes occur from mobile hosts outside of houses, virulence could be reduced.

The second evolutionary process concerns the effects of alternative vertebrate hosts on virulence in humans. The extent of human-bug-human transmission varies substantially geographically; substantial human-bug-human transmission occurring throughout most of *T. cruzi*'s range but is virtually if not entirely absent in the US. Theory and comparative data indicate that vectorborne pathogens should tend to be relatively mild in humans when they rarely cycle in humans (Ewald 1983). About

5-25% of nonhuman vertebrate hosts (racoons and oppossums) in the southern US are infected with *T. cruzi* (Burkholder et al. 1980, Karsten et al. 1992, Pung et al. 1995), and a comparison of such strains with strains from humans in Brazil that they are genetically distinct (Clark & Pung 1994). In the US humans rarely acquire *T. cruzi* via vectors, and appear to be dead-end hosts, probably because of the vector proof housing and low vector densities (Burkholder et al. 1980, Kirchhoff 1993, Barrett et al. 1997). In accordance with theory about the evolution of virulence, such infections appear to be particularly mild in humans, so much so that only three cases of acute Chagas' disease from bug bites had been reported in the US as of 1993 (Woody et al. 1961a, b, Kirchhoff 1993).

This situation is of importance to evolutionary control of *T. cruzi* in countries with endemic Chagas disease because by making houses vector-proof, the importance of human-bug-human cycling relative to enzootic cycling should become greatly reduced, thus causing the evolution of increased specialization of *T. cruzi* on nonhuman vertebrates, reduced specialization on humans, and consequently, reduced virulence in humans. As in the case of malaria, the presence of benign strains could be beneficial through protection against severe strains like a free live vaccine, because benign strains of *T. cruzi* can protect against highly virulent clones (Lauria Pires & Teixeira 1997).

THE MERGING OF EPIDEMIOLOGY WITH EVOLUTIONARY BIOLOGY

The ideas presented above illustrate how the evolutionary considerations of virulence is bringing the health sciences is adding a new dimension to the ideas of the early epidemiologists. After Evandro Chagas deciphered the mode of transmission of *T. cruzi*, he stressed the importance of breaking the domestic cycle of transmission through the vector-proofing of houses. Decades later, in the 1940s, the architects of the mosquito-proofing campaign in Alabama stressed the same point for the control of malaria and demonstrated its utility (Watson 1949). But just as the results of the Alabama study were becoming available, DDT was introduced and successfully used to control malaria in the Mediterranean and South Asia (Harrison 1978). Also at that time the powerful quinine derivatives against malaria were being discovered in response to the cut-off of natural quinine to the Allied powers during World War II. The evidence of the epidemiological value of vector-proofing houses as a control measure against malaria was set aside and largely forgotten in favor of these two more attractive options. A half-century of experience has demonstrated how these two alternatives are incapable of the sort of global eradication that was envisaged at mid-century and have left researchers narrowing their hopes on vaccines as their remaining option for eradication. But broadly effective vaccines have proved elusive; moreover, the evolutionary versatility of plasmodia casts doubt on the long-term success of vaccination – the generation of effective vaccines may be a less formidable challenge than maintaining the efficacy of vaccines after they are put into use. The present therefore seems an opportune time to investigate the possibility of using the evolutionary versatility of plasmodia to our advantage, to generate milder variants.

Like Chagas, by quantifying the frequencies of cholera in areas of London one-and-one-half centuries ago, John Snow demonstrated that the frequencies of cholera were associated with contamination of water supplies (Snow 1855). By integrating evolutionary insights with this kind of epidemiological insight we can add a second dimension to studies of cholera, namely that the harmfulness of pathogens (and hence the harmfulness per infection) is also associated with contamination of water supplies.

Evolutionary considerations strengthen arguments for improving housing and water quality in two ways. First, evolutionary considerations reveal weak spots in programs based on insecticides and antibiotics: the target organisms evolve resistance. Second, evolutionary considerations suggest a previously unrecognized evolutionary benefit of such improvements: the target pathogens should evolve reduced virulence. To evaluate the validity of such evolutionary benefits the suggested interventions (making water supplies pure and houses vector-proof) need to be enacted and studied prospectively in human populations. The molecular and genetic tools are already available or could be readily developed for target pathogens in each category. Molecular determinants of virulence are needed to determine whether evolutionary changes in virulence occur. Molecular phylogenies are needed to categorize pathogens according to their epidemiological history and to structure statistical tests.

Our current state of knowledge already seems sufficient to justify such investments in these interventions ethically and economically. The experimental tracking of epidemiological changes in the frequencies of virulent and mild genotypes should provide conclusive answers to these evolutionary questions while simultaneously providing the epidemiological benefits envisioned by Snow and Chagas.

REFERENCES

Ackerman EA 1956. TVA in its larger setting, p. 244-256. In RC Martin, *TVA. The First Twenty Years. A Staff Report,* University of Alabama Press & University of Tennessee Press.

Barrett VJ, Leiby DA, Odom JL, Otani MM, Rowe JD, Roote JT, Cox KF, Brown KR, Hoiles JA, Saez Alquezar A, Turrens JF 1997. Negligible prevalence of antibodies against *Trypanosoma cruzi* among blood donors in the southeastern United States. *Am J Clin Pathol 108*: 499-503.

Burkholder JE, Allison TC, Kelly VP 1980. *Trypanosoma cruzi* (Chagas) (Protozoa: Kinetoplastida) in invertebrate, reservoir, and human hosts of the lower Rio Grande Valley of Texas. *J Parasitol 66*: 305-311.

CDC - Centers for Disease Control 1996. Dengue fever at the U.S.-Mexico border, 1995-1996. *Morb Mort Weekly Report 45*: 841-844.

Clark CG, Pung OJ 1994. Host specificity of ribosomal DNA variation in sylvatic *Trypanosoma cruzi* from North America. *Mol Biochem Parasitol 66*: 175-179.

Dawson M, Johnson PT, Feldman L, Glover R, Koehler J, Blake P, Toomey KE 1997. Probable locally acquired mosquito-transmitted *Plasmodium vivax* infection - Georgia, 1996 (Reprinted from MMWR, vol 46, p. 264-267, 1997). *JAMA 277*: 1191-1193.

Derryberry OM 1956. Health, p. 193-205. In RC Martin, *TVA. The First Twenty Years. A Staff Report*, University of Alabama Press & University of Tennessee Press.

Elhassan IM, Hviid L, Jakobsen PH, Giha H, Satti GMH, Arnot DE, Jensen JB, Theander TG 1995. High proportion of subclinical *Plasmodium falciparum* infections in an area of seasonal and unstable malaria in Sudan. *Am J Trop Med Hyg 53*: 78-83.

Ewald PW 1983. Host-parasite relations, vectors, and the evolution of disease severity. *Ann Rev Ecol Syst 14*: 465-85.

Ewald PW 1991. Waterborne transmission and the evolution of virulence among gastrointestinal bacteria. *Epidemiol Infect 106*: 83-119.

Ewald PW 1994. *Evolution of Infectious Disease*, Oxford University Press, New York.

Gonzalez JM, Olano V, Vergara J, Arevalo Herrera M, Carrasquilla G, Herrera S, Lopez JA 1997. Unstable, low-level transmission of malaria on the Colombian Pacific Coast. *Ann Trop Med Parasitol 91*: 349-358.

Harrison G 1978. *Mosquitoes, Malaria & Man: A History of the Hostilities Since 1880*, EP Dutton, New York.

Harvey PH, Pagel MD 1991. *The Comparative Method in Evolutionary Biology,* Oxford University Press, Oxford.

Karaolis DKR, Lan R, Reeves PR 1995. The sixth and seventh cholera pandemics are due to independent clones separately derived from environmental, nontoxigenic, non-O1 *Vibrio cholerae*. *J Bacteriol 177*: 3191-3198.

Karsten V, Davis C, Kuhn R 1992. *Trypanosoma cruzi* in wild raccoons and opossums in North-Carolina. *J Parasitol 78*: 547-549.

Kirchhoff LV 1993. Current concepts - American trypanosomiasis (Chagas' disease) - a tropical disease now in the United States. *New Engl J Med 329*: 639-644.

Kun JFJ, SchmidtOtt RJ, Lehman LG, Lell B, Luckner D, Greve B, Matousek P, Kremsner PG 1998. Merozoite surface antigen 1 and 2 genotypes and resetting of *Plasmodium falciparum* in severe and mild malaria in Lambarene, Gabon. *Trans R Soc Trop Med Hyg 92*: 110-114.

Lauria Pires L, Teixeira ARL 1997. Protective effect of exposure to non-virulent *Trypanosoma cruzi* clones on the course of subsequent infections with highly virulent clones in mice. *J Comp Pathol 117*: 119-126.

Lenski RI, May RM 1994. The evolution of virulence in parasites and pathogens: reconciliation between two competing hypotheses. *J Theor Biol 169*: 253-265.

Morens DM, Marchette NJ, Chu MC, Halstead SB 1991. Growth of dengue type-2 virus isolates in human peripheral blood leukocytes correlates with severe and mild dengue disease. *Am J Trop Med Hyg 45*: 644-51.

Pung OJ, Banks CW, Jones DN, Krissinger MW 1995. *Trypanosoma cruzi* in wild raccoons, opossums, and triatomine bugs in southeast Georgia, U.S.A. *J Parasitol 81*: 324-326.

Salles CA, Momen H 1991. Identification of *Vibrio cholerae* by enzyme electrophoresis. *Trans R Soc Trop Med Hyg 85*: 544-547.

Snow J 1855 (1996 reprint). *On the Mode of Communication of Cholera,* 2nd ed., London, Churchill.

Snow RW, Marsh K 1995. Will reducing *Plasmodium falciparum* transmission alter malaria mortality among African children? *Parasitol Today 11*: 188-190.

Tauxe RV, Mintz ED, Quick RE 1995. Epidemic cholera in the new world: translating field epidemiology in now prevention strategies. *Emerg Infect Dis 1*: 141-146.

Vaughn DW, Green S, Kalayanarooj S, Innis BL, Nimmannitya S, Suntayakorn S, Rothman AL, Ennis FA, Nisalak A 1997. Dengue in the early febrile phase: viremia and antibody responses. *J Infect Dis 176*: 322-330.

Watson RB 1949. Location and mosquito-proofing of dwellings, p. 1184-1202. In MF Boyd, *Malariology. A Comprehensive Survey of All Aspects of This Group of Diseases from a Global Standpoint,* Saunders, Philadelphia.

Weber JT, Levine WC, Hopkins DP, Tauxe RV 1994. Cholera in the United States, 1965-1991. Risks at home and abroad. *Arch Intern Med 154*: 551-556.

Weissman JB, Murton KI, Lewis JN, Friedemann CHT, Gangarosa EJ 1974. Impact in the U.S. of the Shiga dysentery pandemic of Central America and Mexico: A review of surveillance data through 1972. *J Infect Dis 129*: 218-23.

Woody NC, Woody HB 1961a. American trypanosomiasis I. Clinical and epidemiological background of Chagas' disease in the U.S. *J Pediat 58*: 568-580.

Woody NC, DeDranous N, Woody HB 1961b. American trypanosomiasis II. Current serologic studies in Chagas' disease. *J Pediat 58*: 738-745.

Wyler DJ 1993. Malaria: overview and update. *Clin Infect Dis 16*: 449-458.

Mem Inst Oswaldo Cruz, Rio de Janeiro, Vol. 93(5): 577-580, Sep./Oct. 1998

Integrated Genetic Epidemiology of Infectious Diseases: The Chagas Model

Michel Tibayrenc

Centre d'Etudes sur le Polymorphisme des Microorganismes (CEPM), UMR CNRS/ORSTOM 9926, ORSTOM, BP 5045, 34032 Monptellier Cedex 01, France

Genetic typing of pathogenic agents and of vectors has known impressive developments in the last 10 years, thanks to the progresses of molecular biology, and to the contribution of the concepts of evolutionary genetics. Moreover, we know more and more on the genetic susceptibility of man to infectious diseases. I propose here to settle a new, synthetic field of research, which I call 'integrated genetic epidemiology of infectious diseases' (IGEID). I aim at evaluating, by an evolutionary genetic approach, the respective impact, on the transmission and pathogenicity of infectious diseases, of the host's, the pathogen's and the vector's genetic diversity, and their possible interactions (co-evolution phenomena). Chagas' disease constitutes a fine model to develop the IGEID methodology, by both field and experimental studies.

Key words: co-evolution - genetic typing - evolutionary genetics - *Trypanosoma cruzi*

Genetic studies dealing with infectious agents, vectors and hosts (for example: genetic susceptibility of man to infectious diseases) have developed until now separately, in a compartmentalized manner. Nevertheless, in an evolutionary point of view, the three actors of infectious disease transmission (the pathogen, the host, and in the case of vector-borne diseases, the vector) have evolved together, and should be considered as the three linked components of a unique phenomenon of co-evolution. When the host evolves (for example, develops specific immune defenses to escape from the damage caused by the pathogen), it shapes in return the evolution and the genetic diversity of the pathogen. It is therefore distressing to analyze separately these three components. I have proposed (Tibayrenc 1998a, b) to settle a new, synthetic field of research, the 'integrated genetic epidemiology of infectious diseases' (IGEID), that will take into account simultaneously the impact, on the transmission and pathogenicity of infectious diseases, of the host's, the pathogen's and the vector's genetic diversity, as well as the interactions (phenomena of co-evolution) of these three parameters. I will advocate here that Chagas' disease constitutes a fine model for throwing the first bases of this ambitious approach.

Fax: +33-4-6741.6299.
E-mail: Michel.Tibayrenc@cepm.mpl.orstom.fr
Received 15 June 1998
Accepted 30 July 1998

WHAT ABOUT THE GENETIC DIVERSITY OF *TRYPANOSOMA CRUZI*?

If we consider the putative impact of the host's, the pathogen's and the vectors' genetic diversity on the transmission and pathogenicity of Chagas' disease, there is little doubt that the best known element is *T. cruzi* genetic variability. Many studies have been published on this theme, and it is possible that *T. cruzi* is one of the pathogenic agents which evolutionary genetics is the best explored. Main results can be briefly summarized as follows: *T. cruzi* natural populations show considerable genetic polymorphism, as revealed by isoenzyme electrophoresis (Miles et al. 1978), kDNA RFLP analysis (Morel et al. 1980) and RAPD (Tibayrenc et al. 1993). The most parsimonious hypothesis to account for this huge genetic polymorphism is that it is the result of long-term clonal evolution with possible occasional bouts of genetic exchange (Tibayrenc et al. 1986, Tibayrenc & Ayala 1988). Recently, these suspected recombinant genotypes have been more precisely characterized as stable hybrid lines, that would propagate clonally after the hybridization event (Bogliolo et al. 1996, Carrasco et al. 1996, Brisse et al. 1998). Among the natural clones of *T. cruzi*, some are widespread and more frequently sampled. They have been given the name of 'major clones' (Tibayrenc & Ayala 1988), since it can be suspected that their epidemiological and pathogenic relevance is considerable. It is most probable that these 'clonal genotypes' identified by a limited set of genetic markers do not correspond to real clones, but rather, tho families of closely related clones. We have pro-

posed (Tibayrenc & Ayala 1991) the term of 'clonet' to refer to sets of stocks that appear identical for a given set of genetic markers in a clonal species. *T. cruzi* clonets are distributed into two main phylogenetic lineages within each of which genetic diversity remains considerable (Tibayrenc 1995, Souto et al. 1996). The second main phylogenetic lineage of *T. cruzi* appears as structured into five lesser subdivisions (Brisse et al. 1998), of which some correspond to either hybrid lines or to formerly identified 'major clones' (Tibayrenc & Ayala 1988) or both. By comparison with other pathogens, *T. cruzi* population structure can be defined as follows: it is a clonal species (Tibayrenc et al. 1986) that is structured into durable genetic subdivisions ('discrete typing units' or DTUs; Tibayrenc 1998a, b). The whole species *T. cruzi* is a DTU, as well as its main and lesser genetic subdivisions. All these DTUs can be characterized by specific genetic markers or 'tags' (Tibayrenc 1998a, b). To some extent, *T. cruzi* DTUs and tags can be equated respectively to monophyletic lineages (clades) and synapomorphic characters, although a strict cladistic approach is difficult here, due to the existence of occasional hybridization events. Still the fact remains that *T. cruzi* overall intraspecific phylogeny appears as robust, considering the strong agreement between the species phylogenies generated by independent sets of genetic markers: isoenzymes and RAPDs (Tibayrenc et al. 1993), and microsatellites (Macedo & Pena, pers. comm.). This striking concordance between three different kinds of genetic markers is clear evidence that the strong genetic distances recorded within *T. cruzi* are due to a real evolutionary divergence rather than to individual genetic diversity within a hypothetical, recent ancestral sexual species, as formerly envisaged (Tibayrenc et al. 1984). It is reasonable to expect that the evolutionary divergence accumulated between *T. cruzi* clonal lineages involves also those genes that govern relevant medical properties such as virulence or resistance to drugs. A possible link between *T. cruzi* genetic variability and Chagas'disease clinical diversity has been suspected by Miles et al. (1981). Montanat et al. (1996) have recently corroborated this hypothesis. Long-term experiments performed in our laboratory show a clear correlation between evolutionary divergence among *T. cruzi* clonal lineages and amount of differences for relevant biological properties such as pathogenicity in mice, *in vitro* drug sensitivity or culture growth speed (Laurent et al. 1997, Pinto et al. 1998, Revollo et al. 1998, De Lana et al. 1998). Certain experiments suggest an interaction between clonal genotypes in artificial mixtures (De Lana et al. 1998). Macedo and Pena (1998) have recently proposed a 'clonal-hystotropic model', which states that *T. cruzi* clonal genotypes infecting the same host have each a specific tropism for given organs. These proposals as well as our results dealing with interactions of clonal genotypes lead to consider that the idea: 'one strain, one pathology' is possibly too simplistic. Still the fact remains that convergent lines of results suggest a profound impact of the phylogenetic diversity of *T. cruzi* natural clones on their relevant biomedical properties.

For studies dealing with the integrated genetic epidemiology, *T. cruzi* constitutes an ideal model, for it is clearly subdivided into clear-cut discrete entities: upper and lesser DTUs, and at a lower level of phylogenetic divergence, the natural clones. The RAPD technique is an abundant source of markers for designing probes and PCR diagnoses specific of either DTUs or natural clones. These specific molecular tools can be conveniently used in the context of integrated genetic epidemiology of Chagas' disease.

THE VECTOR

Although less known than the parasite's genetic diversity, triatomine bugs have been the material for various evolutionary genetic analyses. These studies were based mainly on multilocus enzyme electrophoresis, and have focused either on the intraspecific level (population genetics analysis; Tibayrenc et al. 1981a, b, Dujardin & Tibayrenc 1985, Dujardin et al. 1998) or on between-species comparisons (phylogenetic analysis; Pereira et al. 1996, Solano et al. 1996). These data provide a fine starting basis to include the study of the vector in the integrated genetic epidemiology of Chagas'disease. Nevertheless, it will be necessary to complement isoenzyme typing with more modern molecular tools such as RAPDs or microsatellites, in order to increase the resolution power of triatomine bug genetic characterization.

THE HOST

From the genetic point of view, of the three links of Chagas transmission chain, man is the less known. As a matter of fact, contrary to other parasitic diseases such as malaria or schistosomiasis (Abel & Dessein 1997), nothing is known about possible human genetic susceptibility to Chagas'disease and its different clinical forms. Now the genetic variability of the human species has been widely explored (HLA and microsatellite typing, gene mapping), which should make easier to explore the parameter of host genetic susceptibility in the specific case of Chagas' disease.

INTEGRATED GENETIC EPIDEMIOLOGY OF CHAGAS' DISEASE: EXPERIMENTAL APPROACH

Chagas' disease constitutes a very fine model for experimental studies, since it is possible to establish a complete artificial cycle in the laboratory. The parasite is relatively easy to culture, under epimastigote, trypomastigote and amastigote forms. Rearing the vector is easy too, including through artificial feeding devices, which makes it easier to monitor the experimental parameters (Pinto et al. 1998). Lastly, many mammiferous models (mainly mice) can be used as vertebrate hosts. The principle of an experimental approach of integrated genetic epidemiology is to have only one parameter vary at the same time, while the two other ones are kept as constant as possible. For example, if the impact on Chagas' disease of *T. cruzi* is to be explored (either with pure clonal genotypes or artificial mixtures of genotypes), homogenous triatomine bug and mice strains will be used. When the influence of the vector is explored (both at the level of subspecific and interspecific variability), this will be done, in a given experiment, with only one *T. cruzi* clonal genotype and with a unique mouse strain. Lastly, when the host is considered, various populations of a given strain and various strains (males and females) will be used with the same *T. cruzi* clonal genotype and the same triatomine bug strain. Apart from the empirical observation of the respective impact of the host's, the vector's and the pathogen's genetic diversity on Chagas transmission and pathogenicity, it will be possible to identify the genes that are implied in the infectious process, and to analyze gene regulation phenomena through the analysis of mRNAs with the RNA AP-PCR technique (Welsh et al. 1992). For example, it will be possible to analyse gene expression of given *T. cruzi* clonal genotypes (amastigote, epimastigote and trypomastigote forms) before and after passage through given vector and host populations, or before and after infection of cell cultures, or to compare infected vs non-infected cardiac or digestive cells of dissected mice. Again in these RNA AP-PCR analyses, only one parameter will be allowed to vary at a given time, while the other ones are kept as constant as possible.

FIELD STUDIES

The experimental step is the easiest one to master, and is indispensable. Nevertheless, it definitely has to be completed with a more ambitious approach, which is field studies. This involves the joint analysis of man, triatomine bug and parasite populations. When man is considered, the now well-codified screening with microsatellite markers will have to be used. This makes it possible, through the study of families and control, Chagas-free, populations, to look for possible associations between given parts of the human genome and susceptibility to Chagas' disease and its various clinical forms, through a statistical analysis of linkage disequilibrium. In the same time, isolation of *T. cruzi* stocks from the same populations of patients gives the opportunity to explore possible associations between *T. cruzi* clonal genotypes and clinical forms of Chagas' disease. Lastly, the joint analysis of the genetic variability of triatomine bug populations and of the *T. cruzi* stocks isolated from them could make it possible to increase by far the level of resolution of genetic epidemiological tracking. As a matter of fact, genetic evolution of the vector and of the parasite do not have the same patterns and the same speed, although they are linked. They can give therefore non-redundant, complementary indications on the spread of Chagas'disease epidemics.

CONCLUSION; PERSPECTIVES

IGEID is a very ambitious endeavor, that will need the joint efforts of many different teams having complementary expertises. These various competences are difficult to find in only one country. For these reasons, it is a typical field of research that should be launched in the recently-proposed project of 'European Centre for Control of Infectious Diseases' (ECCID; Tibayrenc 1997a, b). The Chagas model gives the opportunity to launch the first bases of this approach, with the advantages of easy-to-master experimental protocols, and aboundant amount of knowldege on the genetic diversity of the pathogen, and, to a lesser extent, of the vector. The general methodologies developed for Chagas'disease will be applicable to a large extent to other infectious models, especially to the ones that involve related parasites (*Leishmania* and African trypanosomes). The comparative approach advocated for in the case of evolutionary genetics of pathogens (Tibayrenc 1995, 1996) should be retained for IGEID. Indeed, only a comparative IGEID approach will permit to draw the general laws that govern pathogen/host/vector coevolution, and in the same time, to enlighten the specificities of each model.

REFERENCES

Abel L, Dessein AJ 1997. The impact of host genetics on susceptibility to human infectious diseases. *Curr Op Immunol 9*: 509-516.

Bogliolo AR, Lauriapires L, Gibson WC 1996. Polymorphisms in *Trypanosoma cruzi*: evidence of genetic recombination. *Acta Trop 61*: 31-40.

Brisse S, Barnabé C, Tibayrenc M 1998. *Trypanosoma*

cruzi: how many relevant phylogenctic subdivisions are there? *Parasitol Today 14*: 178-179.

Carrasco HJ, Frame IA, Valente AS, Miles MA 1996. Genetic exchange as a possible source of genomic diversity in sylvatic populations of *Trypanosoma cruzi. Am J Trop Med Hyg 54*: 418-424.

De Lana M, Pinto A da S, Barnabé C, Quesney V, Noël S, Tibayrenc M 1998. *Trypanosoma cruzi*: compared vectorial transmissibility of 3 major clonal genotypes by *Triatoma infestans. ExpParasitol,* in press.

Dujardin JP, Tibayrenc M 1985. Etude de 11 enzymes et données de génétique formelle pour 19 loci isoenzymatiques chez *Triatoma infestans* (Hemiptera: Rediviidae). *Ann Soc Belge Méd Trop 65*: 271-280.

Dujardin JP, Schofield CJ, Tibayrenc M 1998. Population structure of Andean *Triatoma infestans*: allozymc frequencies and their epidemiological relevance. *Med Vet Entomol 12*: 20-29.

Laurent JP, Barnabé C, Quesney V, Noël S, Tibayrenc M 1997. Impact of clonal evolution on the biological diversity of *Trypanosoma cruzi. Parasitology 114*: 213-218.

Macedo AM, Pena SDJ 1998. Genetic variability of *Trypanosoma cruzi*: implications for the pathogenesis of Chagas' disease. *Parasitol Today 14*: 119-124.

Miles MA, Povoa M, Prata A, Cedillos RA, De Souza AA, Macedo V 1981. Do radically dissimilar *Trypanosoma cruzi* strains (zymodemes) cause Venezuelan and Brazilian forms of Chagas' disease ? *Lancet 8234*: 1336-1340.

Milcs MA, Souza A, Povoa M, Shaw JJ, Lainson R, Toyé PJ 1978. Isozymic heterogeneity of *Trypanosoma cruzi* in the first autochtonous patients with Chagas' disease in Amazonian Brazil. *Nature 272*: 819-821.

Montanat EE, De Luca GM, Gallerano RH, Sosa R, Blanco A 1996. Characterization of *Trypanosoma cruzi* populations by zymodemes: correlation with clinical pictures. *Am J Trop Med Hyg 55*: 625-628.

Morel CM, Chiari E, Plessmann Camargo E, Mattei DM, Romanha AJ, Simpson L 1980. Strains and clones of *Trypanosoma cruzi* can be characterized by pattern of restriction endonuclease products of kinetoplast DNA minicircles. *Proc Natl Acad Sci USA 77*: 6810-6814.

Pereira J, Dujardin JP, Salvatella R, Tibayrenc M 1996. Enzymatic variability and phylogenetic relatedness among *Triatoma infestans*, *T. platensis*, *T. delpontei* and *T. rubrovaria*. *Heredity 77*: 47-54.

Pinto A da S, de Lana M, Bastrenta B, Barnabé C, Quesney V, Noël S, Tibayrenc M 1998. Compared vectorial transmissibility of pure and mixed clonal genotypes of *Trypanosoma cruzi* in *Triatoma infestans*. *Parasitol Res 84*: 348-353.

Revollo S, Oury B, Laurent JP, Barnabé C, Quesney V, Carrière V, Noël S, Tibayrenc M 1998. *Trypanosoma cruzi*: impact of clonal evolution of the parasite on its biological and medical properties. *Exp Parasitol 89*: 30-39.

Solano P, Dujardin JP, Schofield CJ, Romañan C, Tibayrenc M 1996. Isoenzymes as a tool for *Rhodnius* species identification. *Res Rev Parasitol 56* : 41-47.

Souto RP, Fernandes O, Macedo AM, Campbell DA, Zingales B 1996. DNA markers define two major phylogenetic lineages of *Trypanosoma cruzi*. *Mol Biochem Parasitol 83*: 141-152.

Tibayrenc M 1995. population genetics of parasitic protozoa and other microorganisms. *Adv Parasitol 36*: 47-115.

Tibayrenc M 1997a. European Centres for Disease Control. *Nature* (correspondence) *389*: 433-434.

Tibayrenc M 1997b. Microbes sans frontières and the European CDC. *Parasitol Today 13*: 454.

Tibayrenc M 1998a. Genetic epidemiology of parasitic protozoa and other infectious agents: the need for an integrated approach. *Int J Parasitol 28*: 85-104.

Tibayrenc M 1998b. Beyond strain typing and molecular epidemiology: integrated genetic epidemiology of infectious diseases. *Parasitol Today 14*: 323-329.

Tibayrenc M, Ayala FJ 1988. Isozyme variability of *Trypanosoma cruzi*, the agent of Chagas' disease: genetical, taxonomical and epidemiological significance. *Evolution 42*: 277-292.

Tibayrenc M, Ayala FJ 1991. Towards a population genetics of microorganisms: the clonal theory of parasitic protozoa. *Parasitol Today 7*: 228-232.

Tibayrenc M, Echalar L, Carlier Y 1981a. Comparaison isoenzymatique de deux populations boliviennes (altitude et plaine) de *Triatoma infestans* (Hemiptera Reduviidae). *Cah ORSTOM sér Ent méd Parasitol 19*: 125-127.

Tibayrenc M, Echalar L, Carlier Y 1981b. Données de génétique formelle pour six loci isoenzymatiques chez *Triatoma infestans* (Hemiptera, Reduviidae). *Cah ORSTOM sér Ent méd Parasitol 19*: 121-123.

Tibayrenc M, Neubauer K, Barnabé C, Guerrini F, Sarkeski D, Ayala FJ 1993. Genetic characterization of six parasitic protozoa: parity of random-primer DNA typing and multilocus isoenzyme electrophoresis. *Proc Natl Acad Sci USA 90*: 1335-1339.

Tibayrenc M, Solignac M, Cariou ML, Le Ray D, Desjeux P 1984. Les souches isoenzymatiques de *Trypanosoma cruzi*: origine récente ou ancienne, homogène ou hétérogène? *CR Acad Sci Paris 299*: 195-198.

Tibayrenc M, Ward P, Moya A, Ayala FJ 1986. Natural populations of *Trypanosoma cruzi*, the agent of Chagas' disease, have a complex multiclonal structure. *Proc Nat Acad Sci USA 83*: 115-119.

Welsh J, Chada K, Dalal SS, Cheng R, Ralph D, Mcclelland M 1992. Arbitrarily Primed PCR Fingerprinting of RNA. *Nucl Ac Res 20*: 4965-4970.

Molecular Epidemiologic Typing Systems of Bacterial Pathogens: Current Issues and Perpectives

Marc J Struelens

Service de Microbiologie, Hôpital Erasme and Unité d'Epidémiologie des Maladies Infectieuses, Ecole de Santé Publique, Université Libre de Bruxelles 808, Route de Lennik 1070 Bruxelles, Belgium

The epidemiologic typing of bacterial pathogens can be applied to answer a number of different questions: in case of outbreak, what is the extent and mode of transmission of epidemic clone(s)? In case of long-term surveillance, what is the prevalence over time and the geographic spread of epidemic and endemic clones in the population? A number of molecular typing methods can be used to classify bacteria based on genomic diversity into groups of closely-related isolates (presumed to arise from a common ancestor in the same chain of transmission) and divergent, epidemiologically-unrelated isolates (arising from independent sources of infection). Ribotyping, IS-RFLP fingerprinting, macrorestriction analysis of chromosomal DNA and PCR-fingerprinting using arbitrary sequence or repeat element primers are useful methods for outbreak investigations and regional surveillance. Library typing systems based on multilocus sequence-based analysis and strain-specific probe hybridization schemes are in development for the international surveillance of major pathogens like Mycobacterium tuberculosis. *Accurate epidemiological interpretation of data obtained with molecular typing systems still requires additional research on the evolution rate of polymorphic loci in bacterial pathogens.*

Key words: bacterial typing - DNA polymorphism - pulsed-field gel electrophoresis - ribotyping - RFLP typing - PCR fingerprinting - randomly amplified polymorphic DNA (RAPD) - cross-infection - epidemiology - surveillance

WHY DO WE NEED EPIDEMIOLOGIC TYPING?

Epidemiologic typing systems can be used for *outbreak investigations*, to confirm and delineate the patterns of transmission of one or more epidemic clone(s), to test hypotheses about the sources and vehicles of transmission of these clones and to monitor the reservoirs of epidemic organisms. Typing also contributes to *epidemiologic surveillance* and evaluation of control measures, by documenting the prevalence over time and circulation of epidemic clones in infected populations. Clearly, different requirements will be needed for these distinct applications (Maslow & Mulligan 1996, Struelens et al. 1996).

The basic premise of epidemiologic typing is that isolates of an infectious agent that are part of the same chain of transmission are clonally related, that is the progeny of the same ancestor cell. Extensive genomic and phenotype diversity exists within populations of microbial pathogens of the same species. This diversity reflects the evolutionary divergence arising from mutations and gene flux. Clonally related isolates exhibit significantly more similar characters than unrelated isolates. These distinctive characters, called epidemiological markers, are scored by typing systems which are designed to optimize discrimination between epidemiologically related and unrelated isolates of the pathogen of interest (Maslow & Mulligan 1996, Struelens et al. 1996). The threshold of marker similarity used for definition of a clone need to be adjusted to the species studied, the typing system used, the environmental selective pressure and the time and space scale of the study (Tibayrenc 1995, Struelens et al. 1996). Mutation rate and gene flux vary between species, pathovars and environments. *In vivo* micro-evolution of most pathogens remains poorly understood. Subclonal evolution and emergence of variants that occur in individual hosts or during prolonged transmission can be recognized by several high resolution molecular typing systems, like, for instance, macrorestriction analysis by pulsed-field gel electrophoresis (Struelens et al. 1993, 1996).

CURRENT TECHNOLOGIES: HOW WELL DO THEY FULFILL OUR NEEDS?

In recent years, the development and extensive use of high resolution molecular typing systems based on direct analysis of genomic polymorphism have greatly improved the understanding of the

Fax: +32.555.64.59. E-mail: marc.struelens@ulb.ac.be
Received 15 June 1998
Accepted 30 July 1998

epidemiology of infectious diseases (Maslow & Mulligan 1996, Struelens et al. 1996). However, the rapid diversification and incomplete comparative evaluation of these methods leave the microbiologist and the epidemiologist faced with a number of questions dealing with selection of the appropriate typing system(s) for solving a particular problem, as well as a lack of consensus about interpretation and communication of results.

Several criteria are proposed for evaluating the performance of typing systems (Maslow & Mulligan 1996, Struelens et al. 1996). These criteria include: *typeability, reproducibility, stability, discriminatory power, and epidemiologic concordance. Typeability* refers to the proportion of isolates that can be scored in the typing system and assigned a type, ideally all isolates. *Reproducibility* refers to the ability of the typing system to assign the same type on repeat testing of the same strain. *Stability* is the biological feature of clonally derived isolates to express constant markers over time and generations. The stability of markers may be acceptable even in the presence of variation, provided that the typing system enables recognition of clonal relatedness and does not lead to misclassification of subclonal variants as epidemiologically unrelated. *Discriminatory power* is a key characteristic of typing systems, because it conditions the probability that isolates sharing identical or closely-related types are truly clonal and part of the same chain of transmission. Discriminating power can be calculated based on Simpson's index of diversity. Ideally, the index, based on testing a large number of epidemiologically unrelated isolates, should equal 1. In other words, each independent isolate should be sufficiently different to be assigned to a distinct clone. In practice, a typing system, or combination of systems, displaying a discrimination index greater than 0.95 is acceptable. This level of discrimination corresponds to a 5% probability of erroneously assigning independent isolates to the same clone. *Epidemiologic concordance* is the capacity of a typing system to correctly classify into the same clone all epidemiologically related isolates from a well-described outbreak. Additional comparative studies are needed to establish the relative value of systems currently used for typing microbial pathogens. Moreover, there are important variations in the performance of a given method depending on the species and on modifications of the procedure as applied by different investigators.

In addition to its intrinsic performance when applied to a particular microbial pathogen, a typing system should have practical advantages. Versatility, or the ability to type any pathogen, given minor modifications of the method, is an important advantage for the study of nosocomial infections. Other practical aspects of typing systems include ease of performance and ease of result interpretation, as well as cost and availability of reagents and equipment. Moreover, results should be obtained rapidly enough to be useful in making decisions about management of an outbreak. Infection control problems which require rapid typing data include confirmation that an outbreak is occurring and identification of carriers of the epidemic clone to implement isolation precautions or decolonization therapy. Because there is no optimal typing system that meets all the above requirements, it is as a rule necessary to use a combination of systems. Rapid screening systems can be used initially for preliminary assessment of clonality. Confirmation can be obtained subsequently, if required, by using more reliable but less efficient typing systems. Recent reviews have proposed "optimal" first pass and alternate methods as well guidelines for interpreting differences for a typing number of bacteria when faced with the need to investigate outbreaks (Maslow & Mulligan 1996, Struelens et al. 1996, Tenover et al. 1997).

Methods that index chromosomal DNA polymorphism are the best options for comparative typing of most bacteria, especially nosocomial pathogens (Tenover et al. 1997). Good resolution of genomic restriction fragment length polymorphisms (RFLP) analysis is obtained by: (i) transfer of restriction fragments onto membranes, followed by Southern-blot hybridization with DNA probes, and/ or (ii) use of endonucleases that have infrequent (< 30) recognition sites in the chromosome, followed by separation of these macrorestriction fragments by pulsed-field gel electrophoresis (PFGE). Different types of nucleic acid probes are used for typing: (i) genes encoding metabolic, virulence or resistance functions; (ii) multicopy elements, including insertion sequences and transposons, and (iii) rRNA or rDNA sequences (ribotyping). *Southern blot analysis of gene polymorphism* was found moderately discriminating but highly reproducible and stable. Examples include the *mec* determinant for discriminating strains of methicillin-resistant *Staphylococcus aureus* and the exotoxin A probe for typing *Pseudomonas aeruginosa* strains from cystic fibrosis patients. *IS-fingerprinting*, or Southern blot analyses by using insertion sequences as probes, provides a very reproducible and highly discriminating typing tool. Discrimination is related to the presences of multiple copies of these elements at diverse locations in the chromosome. Careful selection and optimization of probe sequence, restriction endonucleases, electrophoresis and hybridization conditions need to be developed for each species or pathovar to be typed. These

techniques are not rapid and required specialized reagents and expertise. International standardization of technique, reagents, type strains and nomenclature was established by public health reference laboratories for IS6110 RFLP-fingerprinting of *Mycobacterium tuberculosis*, which integrates standard computer analysis of patterns and a common database, and is now widely applied for large scale surveillance of tuberculosis (Bauer et al. 1998).

Ribotyping is the most versatile and the most widely used strategy of Southern blot analysis of bacterial genome polymorphism. The evolutionary conservation of ribosomal RNA makes it applicable as a universal bacterial probe. Many important pathogens, including *Enterobacteriaceae, Listeria, Pseudomonas* sp. and staphylococci have more than five ribosomal operons and thus produce ribotype patterns of 5 to 15 bands. Ribotyping is a robust method that exhibits excellent reproducibility and stability, both *in vitro* and *in vivo* during the course of outbreaks. It is commercially available in a fully automated and well-standardized format. However, its discriminatory power is only moderate, at a level equal or inferior to that of multilocus enzyme electrophoresis. This is related to the fact that ribosomal operons cover less than 0.1% of chromosomal length and tend to cluster in one particular region of the genome. Discrimination of ribotyping depends on species and on choice and number of restriction endonucleases used. No consensus has been achieved on optimal procedure and no general rules are available for interpretation of technically problematic results, like weakly hybridizing fragments.

Macrorestriction analysis resolved by pulsed-field gel electrophoresis has recently emerged as a gold standard for genome fingerprinting of microbial pathogens (Maslow & Muligan 1996, Tenover et al. 1997). Careful selection of low-frequency cleaving enzymes enables cutting the whole bacterial chromosome of any species into less than 30 fragments, typically 10 to 700 kb in size. Periodic change in the orientation of electric field during agarose electrophoresis, or pulsed field gel electrophoresis, allows separation and size determination of these macrorestriction fragments. With minor modifications in the selection of enzymes and "pulsing protocols", PFGE can be applied to any bacterium or yeast. Although direct probing of recognition sequences by rare cutters detects variation in less than 0.01% of the chromosome, large size rearrangements, like sequence duplication, deletion, or insertion, will be readily detected as a shift in fragment size and/or number. In comparison with other typing methods, PFGE has shown equal or greater discriminatory power (Maslow & Muligan 1996). PFGE requires two to four days before results are available and specialized equipment that is more expensive than those required for PCR or Southern hybridization. Nevertheless, because of its superior versatility, reproducibility and resolution, genome macrorestriction analysis is currently a method of choice for typing a majority of nosocomial pathogens and some community-acquired pathogens (Tenover et al. 1997). The sensitivity of PFGE to detect genomic rearrangements makes appropriate interpretation of minor pattern differences a key to its correct application to outbreak investigations and surveillance studies. Interlaboratory standardization has not yet reached a sufficient level to allow the use of common type nomenclature or direct DNA pattern exchange.

In recent years, a number of PCR-based strategies have been developed for strain discrimination of microbial pathogens. In *PCR-gene RFLP typing*, a target sequence, 1 to 2 kb long and known to show polymorphism among strains of the species of interest, is amplified at high stringency. The amplified product is cut with restriction endonucleases and isolates are compared by RFLP pattern. The PCR-serotyping method takes advantages of the conserved sequences at each end of protein antigens genes, like flagellin and outer membrane proteins of Gram-negative pathogens, for amplification of allelic variant sequences encoding the central, antigenically variable portion of these proteins. The polymorphic alleles can be determined by amplicon characterization with suitable restriction endonucleases (PCR-RFLP serotyping (Harrington et al. 1997) or conformational analysis (e.g., single strand conformation analysis, or PCR-SSCP serotyping). The advantages of these methods over conventional serotyping include the unlimited availability of specific reagents, use of universal techniques and typeability of variant strains with cryptic antigens.

Although it is a rapid, simple and reproducible technique, PCR-RFLP typing has shown so far only moderate discrimination. Moreover, it can be biased either by mosaicism due to horizontal transfer, e.g., flagellin gene in *Campylobacter jejuni* (Harrington et al. 1997) or confounded by hypermutation rate at so–called contingency loci that undergo rapid rearrangements in response to environmental changes (e.g., protein A gene polymorphism in *S. aureus* (van Belkum et al. 1996).

As the logical next step, nucleotide *sequencing* of PCR-amplified genes is the most sensitive and accurate means of indexing localized DNA polymorphism for strain typing. However, the time required and cost of the procedure are currently limiting the use of this method which has been

applied to type viruses such as hepatitis viruses and HIV, but also bacteria such as *Streptococcus pyogenes* (Perea Mejia et al. 1997). With the rapid progress of automated, high troughput methods like DNA chip technology (Chee et al. 1996), it is likely that PCR resequencing will be increasingly used for epidemiologic typing of viruses, bacteria and other pathogens in the years to come.

Arbitrarily-primed PCR (AP-PCR) typing, and similar methods like RAPD (random amplified polymorphic DNA) and DAF (DNA amplification fingerprinting), are based on low-stringency PCR amplification by using a single, 10 to 20-mer primer of arbitrary sequence. In the early cycles of the PCR reaction, the primer anneals to multiple sequences with partial homology, and fragments of DNA lying within less than a few kb between annealing sites on opposite DNA strands are amplified. After additional cycles, a strain-specific array of amplified DNA segments of various sizes is obtained. This simple and rapid technique has been successfully applied to genotypic strain delineation and genetic population analysis of a broad range of microbial pathogens, including bacteria, fungi and protozoans. All isolates are typeable and no prior knowledge of target genome sequences is necessary. Discrimination is good and correlates well with other genotyping techniques. The discriminatory power is variable according to number and sequence of arbitrary primers and amplification conditions. In spite of its attractive efficiency, AP-PCR typing suffers from problems in reproducibility and from the lack of consensus rules for interpretation of pattern differences (Maslow & Mulligan 1996, Struelens et al. 1996). A number of technical factors need to be strictly standardized for optimal reproducibility (Grundmann et al. 1997). Progress toward enhanced resolution and reproducibility of analysis of PCR products is achieved by incorporating fluorescent primers in the reaction and performing computer-analysis of amplimer patterns by an automated laser fluorescence detection system. In general, differences in protocols, equipment, or even the batch of reagent used result in different AP-PCR patterns, but the overall clustering and grouping of isolates into identical, similar, or divergent patterns is reproducible. This makes the method adequate for rapid comparative typing but less suitable for library typing in surveillance programs.

Repetitive element PCR (rep-PCR) typing consists of PCR amplification of spacer fragments lying between repeat motifs of the genome by use of two outwardly-directed primers at high stringency. Short, repetitive elements which have been successfully used as targets for rep-PCR typing include the repetitive extragenic palindromes (REPs), the enterobacterial repetitive intergenic consensus (ERIC) sequences, insertion sequences and other species-specific repeat elements (Maslow & Mulligan 1996, Deplano et al. 1997). These rep-PCR strategies produce fewer amplified DNA fragments than AP-PCR, but can nevertheless provide good discriminatory power. Their major advantage is a better reproducibility as compared with AP-PCR analysis, which may enable their standardization for use as library typing systems.

Another set of innovative PCR-based strategy, which also appears to offer high resolution and good reproducibility, are the *amplified fragment length polymorphism (AFLP)* method (Vos et al. 1995) and *infrequent restriction site amplification (IRS-PCR)* (Mazurek et al. 1996). In these methods, a restriction-ligation step produces restricted genomic DNA fragments tagged with specially designed adapters. A set of different primers complementary to these adapters and adjacent nucleotides are then used to PCR amplify various parts of the tagged restriction fragments, thereby selectively highlighting a subset of restriction fragments. More studies are needed to determine the stability of these markers over time, establish criteria for interpretation of pattern differences and evaluate inter-laboratory reproducibility.

Finally, specialized genotyping schemes use reverse *dot blot or line blot binary hybridization patterns* of crude genomic DNA or amplified regions thereof with immobilized, clone-specific DNA probes. This method has been developed for typing of *S. aureus* (van Leeuwen et al. 1996) and *M. tuberculosis* (Kamerbeek et al. 1997). These library probe genotyping systems provide unambiguous, numeric clonal signatures that should be reproducible between laboratories. Inclusion of additional polymorphic sequences should increase the discrimination to the level needed for surveillance of major pathogens. The power of these genotypic hybridization schemes could be much enhanced by the use of high density DNA probe assays, as this technology currently allows parallel analysis of 10^4 target sequences within a few hours.

HOW TO INTERPRETE DIFFERENCES OBSERVED BETWEEN GENOTYPES ?

We use molecular typing systems in epidemiologic studies to determine if isolates are clonally related and thus belong to the same chain of transmission. When a set of isolates show identical DNA banding patterns, this clue to clonality is proportional to the number of typing systems used and their discriminatory power. A problem arises when patterns are similar but not identical. What level of pattern similarity can be used to define clonally or epidemiologically related organisms? This level

needs to be adjusted to the resolving power of the system used, the genomic plasticity of the organism under study and the time scale of the investigation. Suggested rules for interpretation of differences in PFGE patterns, as applied to outbreak investigations (Struelens et al. 1996, Tenover et al. 1997), relate the gradual increase in the number of restriction fragment mismatches with increasing number of genetic differences and with decreasing probability of epidemiologic relatedness. Calculation of restriction/hybridization pattern similarity coefficients and graphical display of pattern relatedness as dendrograms is also useful for interpretation, particularly for large scale studies (Struelens et al. 1996). Altough this quantitative analysis has been criticized as invalid for phylogenetic inferences, because DNA restriction fragment pattern variation is not due to independent events, it is supported by population analysis, e.g. of *P. aeruginosa*. Additional population genetic and micro-evolution studies are needed to provide a better understanding of the nature and frequency of molecular events giving rise to genomic polymorphisms exploited empirically by epidemiologists for strain typing (Tibayrenc 1995, Struelens et al. 1996)

Molecular typing systems are undergoing rapid technical improvements. Advances in the understanding of biological basis of microbial biodiversity at subspecies levels will improve the conceptual framework required for proper epidemiologic interpretation of typing results. Wider application of these systems should shed light to the epidemiology of hospital and community-acquired infections and, therefore, allow for more effective control and prevention strategies.

REFERENCES

Bauer J, Yang Z, Poulsen S, Andersen AB 1998. Results from 5 years of nationwide DNA fingerprinting of *Mycobacterium tuberculosis* complex isolates in a country with a low incidence of *M. tuberculosis* infection. *J Clin Microbiol 36*: 305-308.

Chee M, Yang R, Hubbell E, et al. 1996. Accessing genetic information with high-density DNA arrays. *Science 274*: 610-614.

Deplano A, Vaneechoutte M, Verschraegen G, Struelens MJ 1997. Typing of *Staphylococcus aureus* and *Staphylococcus epidermidis* strains by PCR analysis of inter-IS256 spacer length polymorphisms. *J Clin Microbiol 35*: 2580-2587.

Grundmann HJ, Towner KJ, Dijkshoorn L, et al. 1997. Multicenter study using standardized protocols and reagents for evaluation of reproductibility of PCR-based fingerprinting of *Acinetobacter* spp. *J Clin Microbiol 35*: 3071-3077.

Harrington CS, Thomson-Carter FM, Carter PE 1997. Evidence for recombination in the flagellin locus of *Campylobacter jejuni* : implications for the flagellin gene typing scheme. *J Clin Microbiol 35*: 2386-2392.

Kamerbeek J, Schould L, Kolk A, et al. 1997. Simultaneous detection and strain differentiation of *Mycobacterium tuberculosis* for diagnosis and epidemiology. *J Clin Microbiol 35*: 907-914.

Maslow J, Mulligan ME 1996. Epidemiologic typing systems. *Infect Control Hosp Epidemiol 17*: 595-604.

Mazurek GH, Reddy V, Marston BJ, Haas WH, Crawford JT 1996. DNA fingerprinting by infrequent-restriction-site amplifieation. *J Clin Microbiol 34*: 2386-2390.

Perea Mejia LM, Stockbauer KE, Pan X, Cravioto A, Musser JM 1997. Characterization of group A *Streptococcus* strains recovered from Mexican children with pharyngitis by automated DNA sequencing of virulence-related genes : unexpectedly large variation in the gene (*sic)* encoding a complement-inhibiting protein. *J Clin Microbiol 35*: 3220-3024.

Struelens MJ, and the Members of the European Study Group on Epidemiological Markers (ESGEM), of the European Society for Clinical Microbiology and Infectious Diseases (ESCMID) 1996. Consensus guidelines for appropriate use and evaluation of microbial epidemiologic typing systems. *Clin Microbiol Infect 2*: 2-11.

Struelens MJ, Schwam V, Deplano A, Baran D 1993. Genome macrorestriction analysis of diversity and variability of *Pseudomonas aeruginosa* strains infecting cystic fibrosis patients. *J Clin Microbiol 31*: 2320-2326.

Tenover FC, Arbeit RD, Goering RV, the Molecular Typing Working Group of the Society for Healthcare Epidemiology of America 1997. How to select and interpret molecular typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. *Infect Control Hosp Epidemiol 18*: 426-439.

Tibayrenc M 1995. Population genetics of parasitic protozoa and other microorganisms. *Advances Parasitol 136*: 47-115.

van Belkum A, van Leeuwen W, Kaufmann ME, et al. 1998. Assessment of resolution and intercenter reproducibility of results of genotyping methicillin-resistant *Staphylococcus aureus* by pulsed-field gel electrophoresis of *Sma*I macrorestriction fragments: a multicenter study. *J Clin Microbiol 36*: 1653-1659.

van Belkum A, Riewerts Eriksen N, et al. 1996. Are variable repeats in the *spa* gene suitable targets for epidemiological studies of methicillin-resistant *Staphylococcus aureus* strains? *Eur J Clin Microbiol Infect Dis 15*: 768-769.

van Leeuwen W, Sijmons M, Sluijs J, Verbrugh H, van Belkum A 1996. On the nature and use of randomly amplified DNA from *Staphylococcus aureus*. *J Clin Microbiol 34*: 2770-2777.

Vos P, Hogers R, Bleeker M, et al. 1995. AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Res 23*: 4407-4014.

Vancomycin-resistant Enterococci in Intensive Care Hospital Settings

Daren J Austin/+, Marc JM Bonten*

Wellcome Trust Centre for the Epidemiology of Infectious Diseases, University of Oxford, South Parks Road, Oxford OX1 3PS, UK *Department of Internal Medicine, University Hospital Utrecht, PO Box 85000, 3508, GA Utrecht, Netherlands

Vancomycin-resistant enterococci (VRE) have recently emerged as a nosocomial pathogen and present an increasing threat to the treatment of severely ill patients in intensive-care hospital settings. We outline results of a study of the epidemiology of VRE transmission in ICUs and define a reproductive number R_0; the number of secondary colonization cases induced by a single VRE-colonized patient in a VRE-free ICU, for VRE transmission. For VRE to become endemic requires $R_0 > 1$. We estimate that in the absence of infection control measures R_0 lies in the range 3-4 in defined ICU settings. Once infection control measures are included $R_0 = 0.6$, suggesting that admission of VRE-colonized patients can stabilize endemic VRE.

Key words: vancomycin resistance - transmission dynamics - epidemiology - infection control

Increasing antibiotic resistance in common bacterial pathogens presents a growing threat worldwide. The emergence of vancomycin-resistant enterococci (VRE) as a nosocomial pathogen is a striking example of this new danger to vulnerable patients. In both the United States and the United Kingdon, the frequency with which isolates have been recovered has increased dramatically during the past seven years (CDC 1995, PHLS 1996). For patients infected with VRE treatment options are often limited and control of outbreaks relies heavily on conventional infection control procedures (Husani & Raad 1997).

Molecular epidemiological studies of early endemic infections suggested that single clones were primarily responsible (Boyce et al. 1995). More recently, many outbreaks appear to involve more than one clone, indicating reintroduction (Morris et al. 1995, Slaughter et al. 1996, Bonten et al. 1996). Since enterococci form part of our natural flora, it was believed that new hospital outbreaks were a result of endogenous sources. However recent studies have revealed that transmission of VRE via the hands of transiently colonized heath-care workers (HCWs) is a very important determinant of spread and persistence (Bonten et al. 1996).

DJA thanks the Wellcome Trust for continued support. MJMB thanks Merck Sharpe & Dohme and the Infectious Diseases Society of the Netherlands and Flanders for grant support.
+Corresponding author. Fax: +44-1865-281245. E-mail: daren.austin@zoology.oxford.ac.uk
Received 15 June 1998
Accepted 30 July 1998

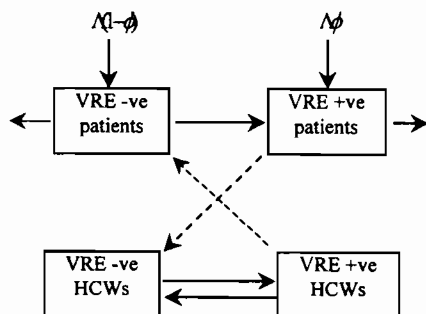
In a previous study conducted at the Cook County Hospital (CCH) Chicago, IL, that measured the relative efficacies of various barrier infection control precautions (Slaughter et al. 1996) rectal cultures were taken daily and environmental cultures monthly (Slaughter et al. 1996). Using pulsed-field gel electrophoresis, a total of 19 strain types of VRE were identified. In a follow-up study it was subsequently confirmed that patients rather than the environment provide the major reservoir of VRE (Bonten et al. 1996).

The transmission dynamics of VRE in an ICU setting can be represented by a set of coupled differential equations with framework summarized by Fig. If we view patients as definitive hosts and HCWs as vectors for transmission, then the structure of the model reduces to that of the Ross-Macdonald equations for malaria transmission (Anderson & May 1991). A central concept in infectious disease transmission is the reproductive number R_0; the number of secondary cases of VRE colonization generated by a single primary case in a VRE-free ICU. If $R_0 > 1$ an outbreak of VRE will persist and become endemic with prevalence $1-1/R_0$. If $R_0 < 1$ the outbreak will fade to extinction. For indirect VRE transmission via HCWs R_0 is defined as

$$R_0 = m\, b_p\, b_s\, r^2 D_p\, D_s$$

where m is the staff-patient ratio, b_p and b_s are the respective probabilities of transmission from HCW-patient and vice-versa, r is the staff-patient contact rate (patient contacts per unit time) and parameters D_s and D_p represent the average duration VRE remains transmissible on the hands of HCWs (typically one hour) and from patients (typi-

cally the duration of their stay in the ICU i.e., days). The contact rate appears as a squared quantity reflecting the patient-HCW-patient nature of transmission.



Model of indirect patient-health care workers (HCW)-patient vancomycin-resistant enterococci (VRE) transmission. Patients are admitted at a rate Λ per day with a fraction ϕ already colonized. Dashed arrows indicate indirect transmission. Colonized patients remain in the ICU for duration D_p and HCWs can transmit VRE for duration D_v.

Infection control measures will influence R_0 in different ways. Barrier precautions such as hand-washing reduce the probability of HCWs transmitting VRE, once colonized, by a factor $(1-p)$ where p is the observed compliance with hand-washing measures. Cohorting HCWs will reduce the effective staff-patient ratio m by a factor $(1-q)$ where q is the proportion of staff cohorted to a single patient. Increasing the numbers of HCWs will increase m but may reduce the patient contact rate r and hence R_0.

Estimates of compliance with barrier precautions have been reported as less than 50% (Doebbbelling et al. 1992), suggesting that VRE will not controlled if $R_0>2$. Cohorting of staff members via one-to-one nursing can give very high cohorting levels, perhaps 80% with a correspondingly higher eradication $R_0<5$. Indeed outbreaks have been brought under control using just such methods (see e.g., Haley et al. 1995).

During the course of a 133-day study of endemic VRE at CCH a mean compliance of 51% was observed and the level of cohorting of HCWs was estimated to be 80%. The mean endemic prevalence of VRE was found to be 36% (95% Cl 3-68) and 15% of patients were already colonized on admission (Slaughter et al. 1996). Our analyses indicate that the effective reproductive number (including infection control measures) $R_0(p,q)=0.6$, corresponding to a true reproductive number $R_0=3.11$ in the absence of infection control measures. We deduce that since $R_0(p,q)<1$, infection control would ordinarily control VRE. However the admission of colonized patients continued to stabilize endemic VRE. The observed reduction in VRE transmission is considerable. In the absence of infection control, the predicted endemic prevalence of VRE is 75% compared with an observed mean of 36%.

The use of molecular epidemiology has demonstrated that the primary determinant of endemic VRE is indirect patient-HCW-patient transmission, rather than environmental or endogenous sources. Using a precise mathematical framework enables careful analysis of the transmission dynamics of VRE and allows for quantitative measurements of both transmission and, more importantly, intervention can be made. As treatment options become more limited, clinicians will become ever more reliant on conventional infection control procedures. The quantitative measurements outlined can be used to assist in enabling better management of limited resources to combat the threat of VRE in ICU hospital settings.

REFERENCES

Anderson RM, May RA 1991. *Infectious Diseases of Humans: Dynamics and Control*, Oxford University Press, 755 pp.

Bonten MJM, Hayden MK, Nathan J, van Voorhis J, Matushek M, Slaughter S, Rice T, Weinstein RA 1996. Epidemiology of colonization of patients and environment with vancomycin-resistant enterococci. *Lancet 348*: 1615-1619.

Boyce JM, Mermal LA, Zervos MJ, Rice LB, Potter-Bynoe G, Giorgio C, Medeiros AA 1995. Controlling vancomycin-resistant enterococci. *Infect Control Hosp Epidemiol 16*: 634-637.

CDC 1995. Recommendations for preventing the spread of vancomycin-resistance. *MMER 44* (RR12): 1-13.

Doebbelling BN, Stanley GL, Sheetz CT, Pfaller MA, Houston AK, Annis L, Li N, Wenzel RP 1992. Comparison of efficacy of alternative hand-washing agents in reducing nosocomial infections in intensive-care units. *NEJM 327*: 88-93.

Haley RW, Cushion NB, Tenover FC, Bannerman TL, Dryer D, Ross J, Sanchez PJ, Siegel JD 1995. Eradication of endemic methicillin-resistant *Staphylococcu aureus* from a neonatal intensive-care unit. *J Infect Dis 171*: 614-624.

Husani R, Raad I 1997. Treatment and prevention of vancomycin-resistant enterococcus. *Curr Opin Intensive Care 10*: 431-434.

Morris Jr JG, Shay DK, Hebden JN 1995. Enterococci resistant to multiple antimicrobial agents, including vancomycin: establishment of endemicity in a University Medical Center. *Ann Internal Med 123*: 250-259.

PHLS 1996. Vancomycin-resistant enterococci in hospitals in the United Kingdom. *CDR Weekly 6*: 1.

Slaughter S, Hayden MK, Nathan C, Hu TC, Rice TJ, van-Voortis J, Matusheka M, Franklin C, Weinstein RA 1996. A comparison of the universal use of gloves and gowns with that of glove use alone on the acquisition of vancomycin-resistant enterococci in a medical intensive-care unit. *Ann Internal Med 125*: 448-456.

Molecular Genetic Analysis of Multi-drug Resistance in Indian Isolates of *Mycobacterium tuberculosis*

Noman Siddiqi/++, Md. Shamim/++, NK Jain*, Ashok Rattan, Amol Amin, VM Katoch***, SK Sharma****, Seyed E Hasnain/+**

National Institute of Immunology, New Delhi, 110067, India *New Delhi T.B. Centre, New Delhi, India **Department of Microbiology, A.I.I.M.S., New Delhi, India ***Central Jalma Institute of Leprosy, Agra, India ****Department of Medicine, A.I.I.M.S., New Delhi, India

A total of 116 isolates from patients attending the out-patient department at the All India Institute of Medical Sciences, New Delhi and the New Delhi Tuberculosis Centre, New Delhi, India were collected. They were analyzed for resistance to drugs prescribed in the treatment for tuberculosis. The drug resistance was initially determined by microbiological techniques. The Bactec 460TB system was employed to determine the type and level of resistance in each isolate. The isolates were further characterized at molecular level. The multi-drug loci corresponding to rpo β, gyr A, kat G *were studied for mutation(s) by the polymerase chain reaction-single strand conformational polymorphism (PCR-SSCP) technique. The SSCP positive samples were sequenced to characterize the mutations in* rpo β, *and* gyr A *loci. While previously reported mutations in the* gyr A *and* rpo β *loci were found to be present, several novel mutations were also scored in the* rpo β *locus.*

Interestingly, analysis of the gyr A *locus showed the presence of point mutation(s) that could not be detected by PCR-SSCP. Furthermore, rifampicin resistance was found to be an important marker for checking multi-drug resistance (MDR) in clinical isolates of* Mycobacterium tuberculosis. *This is the first report on molecular genetic analysis of MDR tuberculosis from India, and highlights the increasing incidence of MDR in the Indian isolates of* M. tuberculosis.

Key words: clinical isolates - *gyr A* gene - multi-drug resistance - *Mycobacterium tuberculosis* - *rpo β* gene - polymerase chain reaction-single strand conformational polymorphism

---

Until recently, the common belief held that tuberculosis (TB) no longer posed a major threat to public health, at least in developed countries. However due to various reasons there is an increasing incidence of TB leading to high morbidity and mortality rates (Bloom & Murray 1992). Furthermore, the association of TB with the AIDS pandemic leading to increase in fatality rates and emergence of multi-drug resistant (MDR) strains of *Mycobacterium tuberculosis* is a cause of grave concern worldwide (Iseman 1994).

Resistance to rifampicin, isoniazid and fluoroquinolones have been well studied and characterized at the molecular level (Honore & Cole 1993, Heym et al. 1993, Miller 1994, Takiff et al. 1994). Rifampicin resistance arises due to mutations in *rpo* β gene encoding the DNA-dependent RNA polymerase. The primary target of rifampicin is the β-subunit of RNA polymerase. The association of the RNA polymerase β (*rpo* β) subunit gene with resistance to rifampicin has been documented previously and subsequent reports from various groups have confirmed this association in clinical isolates of *M. tuberculosis* (Kapur et al. 1994, Williams et al. 1994, Musser 1995, Hasnain et al. 1998). Most of the mutations have been mapped to the 27 codons located at the center of *rpo* β gene that is known to bear mutations that confer rifampicin resistance in *Escherichia coli*. Many of the reported mutations are missense with a few cases of insertions and deletions also (Telenti et al. 1993a, b). Resistance to rifampicin is a relatively rare event and leads to selection of mutants that are already resistant to other components of short-course chemotherapy. Therefore, rifampicin resistance is often regarded as an excellent surrogate marker for MDR-TB (CDC 1993).

Isoniazid acts as the prodrug which is converted to an active form (isonicotinic acid or aldehyde-

---

This project was supported by a grant from the Department of Biotechnology, Ministry of Science and Technology, India.

+Corresponding author. Fax: +9111-616.2125. E-mail: ehtesham@nii.ernet.in

++Both authors contributed equally.

Received 15 June 1998

Accepted 30 July 1998

bearing groups or free radicals) by the kat G-encoded catalase-peroxidase enzyme in *M. tuberculosis*. Isoniazid resistance is due to conversion of Arg463 to Leu in the kat G protein (Heym et al. 1993, 1994). The second mechanism, conferring low level resistance, is mutation in the *inh A* gene which encodes fatty acid synthase. This enzyme requires NADH as a cofactor; the mutant enzyme has been shown to have a lower affinity for NADH and cannot be saturated at NADH concentrations existing within *M. tuberculosis* (Johnsson & Schultz 1994).

The target of fluoroquinolones action is the DNA gyrase, an ATP-dependent type II DNA topoisomerase that catalyses the negative supercoiling of DNA. This enzyme is made up of four units ($\alpha2\beta2$), which are encoded by the *gyrA* and *gyr*B genes respectively. Fluoroquinolones bind to the gyrase and inhibit the supercoiling of DNA. The *gyrA* and *gyrB* genes of *M. tuberculosis* have been cloned and mutations in the quinolone-binding site have been mapped (Takiff et al. 1994).

The present study represents the first report of molecular genetic analysis of clinical isolates of MDR *M. tuberculosis* from India. While association of rifampicin resistance with MDR is evident, we also demonstrate the utility of polymerase chain reaction-single strand conformational polymorphism (PCR-SSCP) for rapidly scoring mutations within the *rpo* β locus. Sequence analysis of the *rpo* β and *gyrA* loci shows that the more common missense mutations are also prevalent in the Indian isolates. This study reaffirms the growing incidence of MDR-TB in India.

MATERIALS AND METHODS

Clinical isolates of *M. tuberculosis* were procured from TB patients attending the out-patient departments at the All India Institute of Medical Sciences and the New Delhi Tuberculosis Center, New Delhi, India. Susceptibility testing of all the *M. tuberculosis* isolates was done by Bactec 460TB system. Minimum inhibitory concentration (MIC) was defined as the lowest drug concentration that inhibited bacterial growth by at least 99%.

DNA was extracted from clinical isolates grown on LJ slants. The colonies were scraped, suspended in TE and subjected to freeze thawing (30 min at -70°C to 100°C for 10 min). This was followed by treatment with lysozyme (40 μg/ml), SDS (0.5%) and proteinase K (50 μg/ml) at 37C for 2 hr. The protein and other contaminants were removed by CTAB precipitation. DNA was finally precipitated with 0.6 volumes of isopropanol. The precipitate was washed twice with 70% ethanol, air dried and re-dissolved in water.

PCR was performed using 150 pmoles of each primer with 1 U of *Taq* polymerase (Bangalore Genei, India), 200 μmoles of each dNTPs and 1.5 mM magnesium chloride. Mutations in the *rpo* β gene, conferring rifampicin resistance were detected with previously reported forward (5 TAC GGT CGG CGA GCT GAT CC 3) and reverse (5 TAC GGC GTT TCG ATG ATG AAC 3) primers. 250 μg of template was amplified in 30 cycles in Perkin Elmer Cetus thermal cycler using the following conditions: 94°C-1min; 56°C-1min; 72°C-2min. The 350 bp *rpo* β amplicon was checked by electrophoresis on a 1.5% agarose gel.

Similar conditions were used for amplification of the *kat G* genes mutational hot spot region using forward (5 GCC CGA GCA ACA CCC 3) and reverse (5 ATG TCC CGC GTC AGG 3) primers.

Sixteen isolates resistant to ofloxacin were checked for mutations in *gyr A*. For *gyr A* amplification the primers used were 5 CAG CTA CAT CGA CTA TGC GA 3 and 5 GGG CTT CGG TGT ACC TCA T 3. The PCR amplification conditions were similar to that of *rpo* β gene amplification except that the annealing temperature was 45°C. The amplicons were purified using Qiaquick PCR purification kit (Qiagen, USA). Samples were initially analyzed for SSCP. Briefly, the samples were heat denatured and electrophoresed on a composite gel (0.25% agarose, 5% acrylamide and 5% glycerol). The gel was then silver stained (BioRad Silver stain Kit) and the DNA bands visualized. A change in the banding pattern as compared to the wild type H37Rv strain was taken as indicative of mutation(s).

Cycle sequencing was performed following the manufacturers protocol using the cyclist *Pfu* Exo-kit (Stratagene, USA).The forward primers of *rpo* β and *gyr A* were used to sequence the SSCP positive samples. The 81 bp and 30 bp sequence corresponding to the hot spot region of the *rpo* β and *gyr A* respectively, were read and compared with the respective sequences of standard H37Rv strain.

RESULTS

The MDR data for the 116 clinical isolates is shown in Table I. A majority of the strains (69%) turned out to be resistant to at least one drug. It is important to mention at the outset that this study has a sampling bias in terms of the drug resistance cases and does not represent the observed frequency of the occurence of drug resistance (~13%) in Indian clinical isolates of *M. tuberculosis* (Ramalingaswami 1998). Between various drugs used in short term chemotherapy for tuberculosis, the isoniazid resistant was most common (56%),

TABLE I

Summary of multi drug resitance in *Mycobacterium tuberculosis* isolates from Indian patients

Drugs	No. of isolates	MRD status	Samples
Isoniazid	65	1 drug	15
Rifampicin	62	2 drugs	25
Ethambutol	27	3 drugs	20
Streptomycin	25	4 drugs	5
Ofloxacin	16	5 drugs	13
None	38		

Total No. of samples: 116

closely followed by rifampicin (53%). The number of isolates resistance to the rest of the drugs was lesser (about 25% for ethambutol and streptomycin) and was least in case of fluoroquinolones. This data corelates well with the treatment regime followed in the TB clinics in India where fluoroquinolones represent the last line of drugs (Pande 1998). We observed that 58 isolates out of 62 positive for rifampicin resistance, were also resistant to atleast one other drug. Therefore, a majority (93%) of the rifampicin resistant strains showed an association with resistance to other drugs thereby supporting earlier observations (Kapur et al. 1994, Williams et al. 1994, Musser 1995, Hasnain et al. 1998) on rifampicin as a surrogate marker for multidrug resistance.

The principle of PCR-SSCP is based on the fact that the two denatured strands of DNA (in this case PCR-amplified) adopt stable intramolecular conformations which may differ from the wild type upon mutation. This causes a change in the electrophoretic mobility of the strands. We utilized SSCP to conduct a primary screening of the *rpo* β, *kat G* and *gyr A* amplicons for the presence of mutations. The results of SSCP analysis reveal that while most of the rifampicin resistant strains did exhibit the expected mobility shifts correlating with point mutations, a very large percentage of isolates resistant to fluoroquinolone and carrying point mutations (revealed upon sequencing), however did not display altered electrophoretic mobility. The results of SSCP analysis of *rpo* β amplicons for few isolates are summarized in Table II. Some typical SSCP gel electrophoresis patterns corresponding to the *rpo* β (Fig. 1a), *kat G* (Fig. 1b) or *gyr A* amplicons (Fig. 1c) are represented.

The precise mutations within the drug resistant loci were identified by direct sequencing of the amplified regions. Sequence analysis (Fig. 2) of the *rpo* β gene hotspot for the rifampicin resistant isolates revealed the presence of many of the common mutations reported earlier (Kapur et al. 1994) in addition to several novel mutations (data not

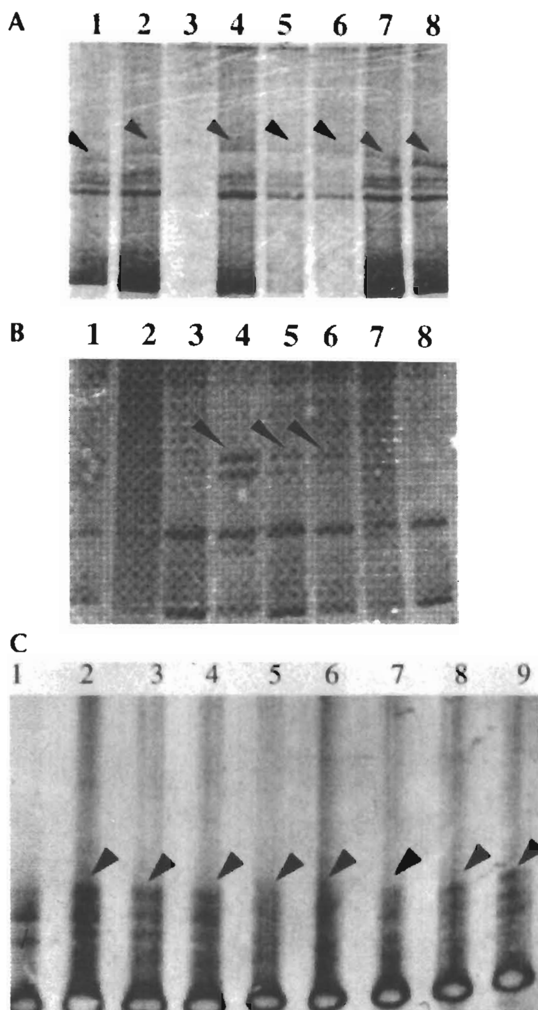


Fig. 1a: PCR-SSCP analysis of representative *rpo β* amplicons.gel picture of few *rpo β* amplicons. Arrow denotes the shift in bands due to the conformational polymorphism as a result of point mutation. Lanes 3 is the control lane; Fig 1b: typical PCR-SSCP pattern displayed by kat G amplicons. Arrowmark denotes the DNA mobility shift due to the conformational polymorphism as a result of point mutation within the kat G locus; Fig. 1c: typical PCR-SSCP gel picture of few mutant *gyr A* amplicons. Arrow mark denotes the DNA mobility shift due to the conformational polymorphism as a result of point mutation within the *gyr A* locus. Lanes 2 to 8 are amplicons from quinolone resistant strains while lane 1 has the wild type H37Rv strain.

shown). Mutations in the *gyr A* gene, however were only of the types already reported earlier (Takiff et al. 1994). The S95T and A90V mutations were the two more common mutations within the Indian isolates of quinolone resistant *M. tuberculosis*. The amino acid changes caused by the corresponding point mutations in the DNA within the *gyr A* loci are shown in Fig. 3.

TABLE II

Susceptibility to rifampicin and polymerase chain reaction-single strand conformational polymophism (PCR-SSCP) results. '?' mark denotes that these isolates were resistant but were not positive in SSCP analysis

Isolate no. analysis	Rifampicin sensitivity (>1 μg/ml)	SSCP
C46	Sensitive	Negative
C49	Resistant	Positive
C56	Resistant	Positive
C64	Resistant	Positive
C68	Sensitive	Negative
C71	Resistant	Negative?
C74	Sensitive	Negative
C78	Resistant	Positive
C80	Sensitive	Negative
C93	Sensitive	Negative
C97	Sensitive	Negative
A1	Sensitive	Negative
A2	Sensitive	Negative
A3	Sensitive	Negative
A4	Sensitive	Negative
A5	Sensitive	Negative
A6	Sensitive	Negative
A7	Sensitive	Negative
A8	Sensitive	Negative
A9	Sensitive	Negative
B1	Sensitive	Negative
B2	Sensitive	Negative
B3	Sensitive	Negative
B4	Sensitive	Negative
B5	Sensitive	Negative
B6	Resistant	Positive
B7	Sensitive	Negative
F1	Resistant	Positive
F2	Resistant	Positive
F3	Resistant	Positive
F4	Resistant	Positive
F5	Resistant	Positive
F6	Resistant	Positive
F7	Resistant	Positive
F8	Resistant	Positive
F9	Resistant	Positive
D1	Sensitive	Negative
D2	Resistant	Negative?
D3	Resistant	Negative?
D4	Resistant	Negative?
D6	Resistant	Negative?
D7	Resistant	Positive
D9	Resistant	Positive
D10	Resistant	Positive
D11	Resistant	Positive
D12	Resistant	Positive
D18	Sensitive	Negative
V1	Resistant	Positive
V2	Resistant	Positive
V3	Resistant	Positive
V4	Resistant	Positive
V5	Resistant	Positive
V6	Resistant	Positive
V7	Sensitive	Negative
V8	Resistant	Positive
V9	Resistant	Positive
V10	Resistant	Positive
V11	Resistant	Positive
V12	Sensitive	Negative
V13	Resistant	Positive
V14	Resistant	Positive
V15	Resistant	Positive
V16	Resistant	Positive
V17	Resistant	Positive
V18	Resistant	. Positive

DISCUSSION

In this study we present molecular genetic analysis of rifampicin, isoniazid and quinolone resistance in Indian clinical isolates of MDR *M. tuberculosis*. Point mutations within *rpo* β , *kat G* and *gyr A* genes respectively lead to amino acid polymorphism in the target protein of the drug resulting in drug resistance (Rattan et al. 1998). MDR does not appear to arise due to the acquisition of a transposable element or a plasmid carrying drug resistant marker, but is perhaps a reflection of stepwise acquisition of new mutations in the genes for different drug targets. Alterations in the chromosomal genes are random but get selected due to poor compliance or prescription. Inadequate prescription of chemotherapy, poor compliance of the drug regime and in recent times infection with HIV have caused an increase in the selection of MDR strains of *M. tuberculosis* (Lederberg 1998). A lack of monitoring programs and poor follow up of the patients health has caused an increase in relapse cases for TB. In most cases, the secondary infection is by drug resistant mycobacterium. What is of graver concern is the increasing incidence of primary infection by MDR- *Mycobacterium tuberculosis* (Hasnain et al. unpub. data). Patients infected with a rifampicin resistant strain of *M. tuberculosis* generally have a poor prognosis, particularly because rifampicin resistance is often associated with resistance to other frontline drugs.

We found a majority of the rifampicin resistant Indian isolates to be resistant to at least one another anti-tubercular drug, supporting the idea of using rifampicin resistance as a surrogate marker for MDR TB. PCR-SSCP analysis while offering a rapid method for detecting MDR particularly for rifampicin resistance however, has limitations as it failed to detect mutations within the *gyr A* locus. These results representing a first report on the molecular genetic analysis of MDR in clinical isolates of *M. tuberculosis* from India have important bearing on the management and control of this reemerging infectious disease.

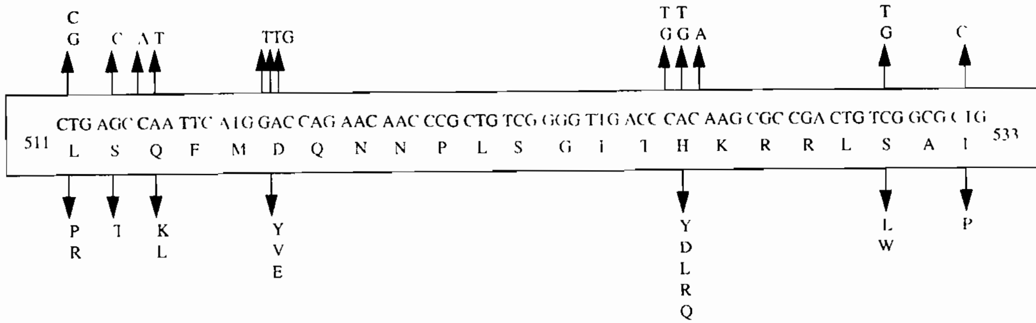


Fig. 2. common mutations in the *rpo β* gene. Top panel shows the nucleotide changes that are similar to previously reported mutations. The bottom panel depicts the corresponding amino acid changes within the RNA polymerase *β* subunit.

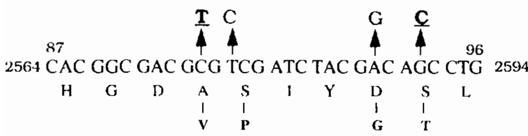


Fig 3: mutations within the *gyr A* gene. The top panel shows the nucleotide change while the bottom panel depicts the corresponding amino acid changes. The common mutations in the Indian isolates are depicted in bold letters.

ACKNOWLEDGMENTS

To Mr Sunder Singh Bisht for his help in sequencing and SSCP.

REFERENCES

Bloom BR, Murray JL 1992. Tuberculosis: commentary on a reemergent killer. *Science 257*: 1055-1064.

CDC - Centers for Disease Control 1993. Initial therapy for tuberculosis in the era of multi-drug resistance. recommendations of the advisory council for the elimination of tuberculosis. *MMWR 42*: RR-7.

Hasnain SE, Amin A, Siddiqi N, Shamim M, Jain NK, Rattan A, Katoch VM, Sharma SK 1998. Molecular genetics of multiple drug resistance (MDR) in *Mycobacterium tuberculosis*, p. 35-40. In RLSinghal & OP Sood (eds), *Drug Resistance: Mechanism and Management*, Proceedings of the Fourth Annual Ranbaxy Science Foundation Symposium Communicore Publishers, New Delhi, India

Heym B, Alzari PM, Honore N, Cole ST 1994. Missense mutations in the catalase-peroxidase gene, katG, are associated with isoniazid resistance in *Mycobacterium tuberculosis. Mol Microbiol 15*: 235-245.

Heym B, Zhang Y, Poulet S, Young D, Cole ST 1993. Characterization of the katG gene encoding a catalase-peroxidase required for the isoniazid susceptibility of *Mycobacterium tuberculosis. J Bacteriol 175*: 4255-4259.

Honore N, Cole ST 1993. The molecular basis of rifampin resistance in *Mycobacterium leprae. Antimicrob Agents Chemother 37*: 414-418.

Iseman MD 1994. Evolution of drug-resistant tuberculosis: A tale of two species. *Proc Natl Acad Sci USA 91*: 2428-2429.

Johnsson K, Schultz PG 1994. Mechanistic studies of the oxidation of isoniazid by the catalase peroxidase from *Mycobacterium tuberculosis. J Am Chem Soc 116*: 7425-7426.

Kapur V, Ling Ling Ll, Iordanescue S, Krieswirth BN, Musser JM 1994. Characterization by automated DNA sequencing of mutations in the gene (*rpo* β) encoding the RNA polymerase β subunit in rifampin *Mycobacterium tuberculosis* strains from New York and Texas. *J Clin Microbiol 32*: 1095-1098.

Lederberg J 1998. The future of infectious disease, p.5-14. In RLSinghal & OP Sood (eds), *Drug Resistance: Mechanism and Management*, Proceedings of the Fourth Annual Ranbaxy Science Foundation Symposium Communicore Publishers, New Delhi, India.

Miller LP, Crawford JT, Shinnick TM 1994. The *rpo* β gene of *Mycobacterium tuberculosis. Antimicrob Agents Chemother 38*: 805-811.

Musser JM 1995. Antimicrobial agent resistance in mycobacteria: molecular genetic insights. *Clin Microbiol Rev 8*: 496-514.

Pande JN 1998. Multidrug resistant tuberculosis: Current concepts and future directions for management, p. 87-90. In RLSinghal & OP Sood (eds), *Drug Resistance: Mechanism and Management*, Proceedings of the Fourth Annual Ranbaxy Science Foundation Symposium Communicore Publishers, New Delhi, India.

Ramalingaswami V 1998. Opening address, p. 1-3. In RLSinghal & OP Sood (eds), *Drug Resistance: Mechanism and Management*, Proceedings of the Fourth Annual Ranbaxy Science Foundation Symposium Communicore Publishers, New Delhi, India.

Rattan A, Kalia A, Ahmad N 1998. Multidrug resistant tuberculosis: Molecular perspectives. *Emerg Infec Dis 4*: 195-209.

Takiff HE, Salazar L, Guerrero C, Philipp W, Huang WM, Kreiswirth B, Cole ST, Jacobs Jr WR, Telenti A 1994. Cloning and nucleotide sequence of the *Mycobacterium tuberculosis gyr*A and *gyr*B genes, and characterization of quinolone resistance mutations. *Antimicrob Agents Chemother 38*: 773-780.

Telenti A, Imboden P, Marchesi F, Lowrie D, Cole S, Bodmer T 1993a. Detection of rifampicin-resistance mutations in *M. tuberculosis*. *Lancet 341*: 647-650.

Telenti A, Imboden P, Marchesi F, Schmidheini T, Bodmer T 1993b. Direct automated detection of rifampicin-resistant *Mycobacterium tuberculosis* by polymerase chain reaction and single-strand conformation polymorphism analysis. *Antimicrob Agents Chemother 37*: 2054-2058.

Williams DL, Wanguespack C, Eisenach KD, Bates JH, Crawford JT 1994. Characterization of rifampin resistance in pathogenic mycobacteria. *Antimicrob Agents Chemother 38*: 2380-2386.

Molecular Basis of Ribotype Variation in the Seventh Pandemic Clone and its O139 Variant of *Vibrio cholerae*

Ruiting Lan, Peter R Reeves/+

Department of Microbiology, University of Sydney, NSW 2006, Sidney, Australia

Ribotyping has been widely used to characterise the seventh pandemic clone including South American and O139 variants which appeared in 1991 and 1992 respectively. To reveal the molecular basis of ribotype variation we analysed the rrn *operons and their flanking regions. All but one variation detected by* Bgl*I, the most discriminatory enzyme, was found to be due to changes within the* rrn *operons, resulting from recombination between operons. The recombinants are detected because of the presence of a* Bgl*I site in the 16S gene in three of the nine* rrn *operons and/or changes of intergenic spacer types of which four variants were identified. As the frequency of* rrn *recombination is high, ribotyping becomes a less useful tool for evolutionary studies and long term monitoring of the pathogenic clones of* Vibrio cholerae *as variation could undergo precise reversion by the same recombination event.*

Key words: ribotyping - *rrn* recombination - seventh pandemic - O139 - *Vibrio cholerae*

Vibrio cholerae is the agent responsible for cholera which was first described in 1854. However, the natural habitat of *V. cholerae* is the aquatic environment. Environmental *V. cholerae* are diverse and most are nontoxigenic. There are more than 190 O antigens identified (Yamai et al. 1997). The best known forms are O1 and recently O139. Both cause cholera, currently mostly in developing countries. Seven pandemics of cholera are recognised since 1817. The seventh started in 1961 and continues to the present day.

The seventh pandemic clone has been studied by various molecular methods. Ribotyping, a form of restriction fragment length polymorphism analysis using rRNA genes, was shown to be very discriminatory in revealing variation. rRNA sequences are highly conserved and the genes (*rrn*) are present as multiple copies in the genome of many bacteria. The typing therefore provides information on several flanking regions simultaneously. There have been quite a few studies on the epidemiology and molecular typing of *V. cholerae.* Koblavi et al. (1990) were the first to employ ribotyping to fingerprint *V. cholerae* strains and Popovic et al. (1993) proposed a standardised scheme for typing *V. cholera* using *Bgl*I restriction enzyme to allow public laboratories to follow the movement and identify the origins of *V. cholerae* strains. Ribotyping has since been widely used to characterise the sixth pandemic clone (Faruque et al. 1993) and the seventh pandemic clone (Karaolis et al. 1994, Faruque et al. 1995), the South American and O139 variants (Wachsmuth et al. 1993, Popovic et al. 1995, Dalsgaard et al. 1997) and other O1 outbreaks (Coelho et al. 1995). Ribotype variation has been found in the South American isolates in the four years of its spread from 1991 (Dalsgaard et al. 1997) and in the O139 variant which appeared in 1992 (Popovic et al. 1995, Faruque et al. 1997).

As the *rrn* operons are conserved, variation detected in ribotyping has been generally assumed to be due to variation in flanking regions. We studied the changes behind ribotype variation to help us understand the evolution of pandemic clones (Lan & Reeves 1998).

RIBOTYPE VARIATION IN THE SEVENTH PANDEMIC CLONE

The seventh pandemic clone is an very interesting clone to study because accurate dates of development are known. In our ribotyping study with a total of 47 strains isolated from 1961 (Indonesia and Hong Kong) to 1993 including African and Asian isolates (Karaolis et al. 1994). *Bgl*I detects most of the polymorphisms. The *Bgl*I ribotypes are summarised in Table I. There are 11 ribotypes. Ribotype G was present at the start of the pandemic and 14 type G strains are all from Asia. In the first 10 years of spread there was no detectable variation. In fact after 1966 there was a lull period with relatively little cholera. Ribotype H appeared in 1970 in Asia and spread to Africa;

This work was supported by a grant from the National Health and Medical Research Council of Australia.
+Corresponding author. Fax: +612 9351.4571. E-mail: reeves@angis.usyd.edu.au
Received 15 June 1998
Accepted 30 July 1998

Africa being free from cholera before 1970. The resurgence in 1970 seems to be associated with a substantial increase in genetic variation. There are other minor ribotypes with one or two isolates from Asia and Africa.

TABLE 1
Summary of *Bgl*I ribotypes detected in the seventh pandemic clone (Karaolis et al. 1994)

Ribotype	No. isolates	Year of isolation	Location of isolation
G	14	1961-1991	Asian region only
H	22	1970-1993	Asia and Africa
I	1	1971	Burma
J	2	1971, 1974	Chad
K	1	1972	Senegal
L	1	1978	Malaysia
M	1	1988	Zaire
N	2	1989, 1992	HK & Indonesia
O	1	1990	Malawi
P	1	1991	India
Q	1	1993	Indonesia

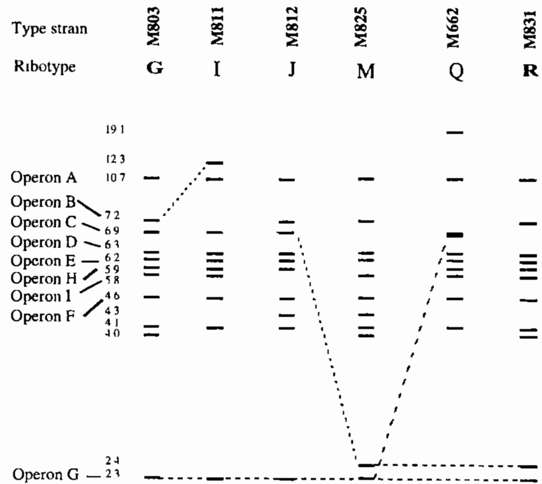
rrn OPERON FEATURES IN THE EARLY SEVENTH PANDEMIC ISOLATE M803

Majumder et al. (1996) mapped seven operons to the genome of *V. cholerae* strain 569B. However, ribotyping in *V. cholerae* detects 10 or more bands in *Bgl*I digests in the seventh pandemic isolates (Karaolis et al. 1994) and it was not clear how to assemble seven operons from the patterns. We know that in general a *rrn* operon has three genes in the order 16S-23S-5S and there is a unique I-*Ceu* I site in the 23S gene (Liu & Sanderson 1995). We used 2 DNA probes, one specific to the 5' region of the 16S gene and the other specific to the 3' region of the 23S gene, and an oligo probe from 16S 5' end to probe Southern blots of *Bgl*I digests. From the probings we concluded that there are nine operons in the strain of M803, an early seventh pandemic isolate. Three operons have a *Bgl*I in their 16S gene. We also used the 23S probe to probe strain M803 DNA digested with I-*Ceu* I and *Bgl*I, which showed only a strong 1 kb band rather nine bands as expected and thereby discovered that there is a *Bgl*I site near the end of the 23S gene. However, the number of operons is different from the seven operons detected by Majumder et al. (1996). Later we discovered that two of the nine operons are in tandem with another operon. Thus there are only seven loci on the chromosome. A recent study using pulsed-field gel electrophoresis of I-*ceu* I digest from the Majumder group (Nandi et al. 1997) found that there are nine fragments, one of which is 6 kb, the equivalent of a tandem operon. However, we could not reconcile our data with their conclusion of eight independent *rrn* loci.

IDENTIFICATION OF ALTERNATIVE FRAGMENTS FOR EACH RIBOTYPE CHANGES BY OPERON SPECIFIC PROBING

Due to the presence of a *Bgl*I site at the end of the 23S genes, changes in the 23S proximal DNA are not detected by ribotyping and all *Bgl*I variation involves the 16S genes or their flanking regions. PCR walking was used to obtain sequence immediately upstream of the 16S genes. We then made operon specific probes and probed *Bgl*I digests of different ribotypes to analyse the variation. The probing experiments are summarised in the Fig. with schematic representation of each *Bgl*I change.

For operons B, C, and G, alternative bands were identified through probing. For operon B, the alternative bands are 7.2 kb, 12.3 kb and 19.1 kb. We found that these alternative bands result from loss of the *Bgl*I site in the 16S gene.The 7.2 kb band has only 5' end of the 16S gene because of the *Bgl*I site in the 16S gene. We sequenced the region around the *Bgl*I site in M803, M811, and M662 of operon B. A single base difference from C to T at base 849 determines the presence or absence of the elevenmer *Bgl*I recognition sequence. However the change to 19.1 kb in M662 from 7.2 kb in M803 or 12.3 kb in M811 also involves



Schematic representation of *Bgl*I ribotype data of *Vibrio cholerae* seventh pandemic isolates to indicate variation in specific operons among ribotypes G, I, J, M and Q and the O139 variant R. The operons are indicated to the left for each band representing the 5' 16S gene. Alternative bands of the same operon are indicated by a dotted line for operons B, C, and G. The size of each band is indicated on the left. After Karaolis et al. (1994) and Lan and Reeves (1998).

change in the flanking region of the operon. We have yet to walk to that *Bgl*I site. For operon C operon-specifc probing showed that the 2.4 kb band in M825 is a replacement of the 6.9 kb fragment which is due to the creation of a *Bgl*I site in the 16S gene. For operon G the alternative band for the 2.3 kb band is similar in size to the 6.9 kb operon C fragment. Apparently this resulted in the newly created band being masked. This change is due to the loss of *Bgl*I site in the 16S gene.

We found that operons H and I are tandem operons and absence of a band means deletion of the operon. We did Southern probing of I-*ceu* I digests to show that H and I are immediately downstream of other *rrn* operons. For tandem operons I-*Ceu* I digestion, which cuts uniquely in 23S gene, would release a fragment from the I-*Ceu* I site of the upstream operon to the I-*Ceu* I site of the downstream operon.

SIZE VARIATION IN INTERGENIC SPACERS

The bands of 4.0, 4.1 and 4.3 kb in the Fig. are fragments having part of 16S gene and the whole of 23S gene and variation in size of these bands is due to size differences of the intergenic spacer. We analysed the intergenic spacer of the three operons B, C and G. The spacer region was amplified by nested PCR. Three types of spacers were initially found: small (S) 411 bp, large (L) 692 bp, and medium (M1) 488 bp. We later identified another spacer variant (M2) 587 bp.

REASSORTMENT OF SPACER TYPES AND *rrn* RECOMBINATION

The nature of variation detected by *Bgl*I ribotyping is summarised in Table II, including eleven ribotypes of the seventh pandemic and its O139 variant. It is evident that there was extensive reassortment of spacers. For operon B, there are four forms: S, *Bgl*I-; L, *Bgl*I+; L, *Bgl*I-; and M2, *Bgl*I- generated from the S, *Bgl*I+ form of the early strains. Both operons C and G have two additional forms. So there are in total eight types of changes to *rrn* operons which we attribute to homologous recombination between operons. We don't know why the most frequent changes are in operon B. The other frequent change is deletion of operon I. Deletions might be due to single or multiple events. The majority of the ribotypes seems to have arisen from ribotype G by a single recombination event though M, N and Q require two or more recombinational changes.

RIBOTYPE VARIATION IN THE O139 VARIANT

It is well recognised from multilocus enzyme electrophoresis (MLEE) and other data that the O139 Bengal strain is derived from the seventh pandemic strain. A single ribotype, R, (represented by strain M831) was identified in our previous study (Karaolis et al. 1994). The O139 clone is very similar to the Asia-dominant ribotype G. The only difference in the *Bgl*I digests of ribotypes R and G is that in ribotype R the 6.9 kb operon C fragment

TABLE II

Spacer types and status of *Bgl*I site in the 16S RNA gene for operons B, C and G and status of operons H and I in the ribotypes (G-Q) of the seventh pandemic clone and ribotype R of the O139 variant

		Operon B		Operon C		Operon G		Operon	Operon
Ribotype	Strain	Spacer [a]	*Bgl*I site [b]	Spacer	*Bgl*I site	Spacer	*Bgl*I site	H	I
G	M803	S	+	M1	-	M1	+	Present	Present
H	M807	. [c]	.	.	.	.	.	.	Del
I	M811	L	-	.	.	.	.	.	.
J	M812	L	.	.	.	.	.	.	Del
K	M813	.	.	S	+	.	.	.	del
L	M820	L	-	.	.	.	.	.	Del
M	M825	L	.	.	+	S	.	Del [d]	.
N	M799	M2	-	.	.	.	-	.	.
O	M826	L	.	.	.	.	.	Del	Del
P	M654	.	-	.	.	.	.	.	Del
Q	M662	M2	-	.	.	.	-	.	.
R	M831	.	.	.	+	.	.	.	Del

a: the spacer types are small (S) 431bp, large (L) 711bp, medium (M1) 509bp and a variant of size between M1 and L (M2) 607bp; *b*: +/- indicates presence or absence of *Bgl*I site at base 838 of the 16S RNA gene; *c*: status identical to ribotype G is indicated by a dot; *d*: deletion of an operon.

is replaced by a 2.4 kb band due to the gain of a *BglI* site in the 16S gene. It is very interesting that Popovic *et al* (1995) detected two ribotypes, 3a and 5a, in O139 isolates. Ribotype 3a is identical to ribotype R. Ribotype 5a differs from ribotype 3a by the absence of operon I. Popovic *et al* (1995) noted that ribotypes 3a and 5a are very similar to their ribotypes 3 and 5 of seventh pandemic clone respectively, which implied that ribotype 3a may be derived from ribotype 3 and 5a derived from 5. Most likely ribotype 5a is derived from ribotype 3a through an independent operon I deletion rather than from ribotype 5. If 3a and 5a were derived separately, it would require independent transfer of a new O antigen gene cluster. This also illustrates the weakness of ribotyping for defining relationships between strains.

RIBOTYPE VARIATION IN THE SOUTH AMERICAN VARIANT

In our ribotyping study (Karaolis et al. 1994), only one South American isolate was included which belongs to ribotype H. Dalsgaard et al. (1997) analysed 50 South American isolates isolated in Peru from 1991 to 1995 and found four ribotypes: R1, R2, R3 and R4. The publication of good quality Southern blot by Dalsgaard et al. (1997) and use of probes identical to our study allow us to interpret their results in terms of the basis of ribotype changes. R1 is identical to our ribotype H. R2 differs from R1 by the appearance of a 4.3 kb band and disappearance of a 4.0 kb band representing the change in spacer type from small to large in operon C by *rrn* recombination. R3 differs from R1 in the presence of a 5.6kb band which is equivalent to the position of operon I and the pattern is identical to our ribotype G, the frequent ribotype isolated in Asian region only. It seems that a R1 (ribotype H pattern) strain reversed back to ribotype G pattern by recreation of the tandem operon I, presumably also due to *rrn* recombination. This is an good example where identical ribotypes may not mean genetic similarity. It is possible that the R3 strain is actually a ribotype G strain separately introduced into South America and R3 is not derived from R1. The early South American isolates differ at the locus of leucine aminopeptidase in MLEE from other seventh pandemic isolates. Therefore the allele profile could be used to determine whether R3 is derived from R1. R4 is very different from the other three ribotypes. In our opinion R4 is not developed from R1.

GENERAL CONCLUSIONS

It has always been assumed that ribotype variation is due to changes outside the *rrn* operons. This study shows that ribotyping detects two types of changes of very different nature: within operon changes produced by *rrn* recombination or mutational changes outside the operons. For the seventh pandemic clone the varaition observed is predominantly *rrn* recombination. The frequency of *rrn* recombination is high. Ten new ribotypes were found in 46 isolates over 33 year span in our previous study (Karaolis et al. 1994). Nine of the ten ribotypes were generated by *rrn* recombination. There are new ribotypes found in the South American isolates in its first four years spread (Dalsgaard et al. 1997). Similarly new ribotypes were found in the O139 variant (Popovic et al. 1995, Faruque et al. 1997).Therefore, the level of parallel and reversal changes will also be high. And similar ribotype does not necessarily reflect genetic similarity. Thus ribotyping is not suitable for long term monitoring of the seventh pandemic clone or any *V. cholerae* clones. Other species need to be studied to see whether this is a general phenomenon.

However, some laboratories may wish to continue to use *BglI* ribotyping for typing seventh pandemic isolates as an accessary tool as it has been well established technically. In such case we recommend using a fragment from the 16S gene encompassing the *BglI* site as a probe, eg the 1 kb fragment from base 21 to base 1097 of the 16S gene in this study, which will produce bands of uniform intensity and identical patterns to probing using a mixture of 16S and 23S rRNA or rDNA (Lan & Reeves 1998). Using a 16S and 23S rRNA as probe the hybridisation signal for those fragments having only part of the 16S gene is much weaker than for other fragments which may lead to misidentification of patterns. The use of probes including 5S gene or any flanking sequences in addition to 16S and 23S genes is not recommended as it produces ribotyping patterns not comparable to those of other laboratories. Separation of the four bands between 6.3 kb to 5.8 kb, and the three bands between 4.0 kb to 4.3 kb is usually poor. Reference ribotypes representing each band variation should be included for comparison to determine the present or absent of a band to increase accuracy.

REFERENCES

Coelho A, Andrade JRC, Vincente ACP, Salles CA. 1995. New variant of *Vibrio cholerae* O1 from clinical isolates in Amazonia. *J Clin Microbiol 33*: 114-118.

Dalsgaard A, Skov MN, Serichantalergs O, Echeverria P, Meza R, Taylor DN 1997. Molecular evolution of *Vibrio cholerae* O1 strains isolated in Lima, Peru, from 1991 to 1995. *J Clin Microbiol 35*: 1151-1156.

Faruque SM, Ahmed KM, Siddique AK, Zaman K, Alim ARMA, Albert MJ 1997. Molecular analysis of toxigenic *Vibrio cholerae* O139 bengal strains isolated in Bangladesh between 1993 and 1996: evidence for emergence of a new clone of the bengal Vibrios. *J Clin Microbiol 35*: 2299-2306.

Faruque SM, Alim ARMA, Rahman MM, Siddique AK, Sack RB, Albert MJ 1993. Clonal relationships among Classical *Vibrio cholerae* O1 strains isolated between 1961 and 1992 in Bangladesh. *J Clin Microbiol 31*: 2513-2516.

Faruque SM, Roy SK, Alim ARMA, Siddique AK, Albert MJ 1995. Molecular epidemiology of toxigenic *Vibrio cholerae* in Bangladesh studies by numerical analysis of rRNA gene restriction patterns. *J Clin Microbiol 33*: 2833-2838.

Karaolis DKR, Lan R, Reeves PR 1994. Molecular evolution of the 7th pandemic clone of *Vibrio cholerae* and its relationship to other pandemic and epidemic *V. cholerae* isolates. *J Bacteriol 176*: 6199-6206.

Koblavi S, Grimont F, Grimont PAD 1990. Clonal diversity of *Vibrio cholerae* O1 evidenced by rRNA gene restriction patterns. *Res Microbiol 141*: 645-657.

Lan R, Reeves PR 1998. Recombination between rRNA operons created most of the ribotype variation observed in the seventh pandemic clone of *Vibrio cholerae*. *Microbiology 144*: 1213-1221.

Liu S-L, Sanderson KE 1995. Rearrangements in the genome of the bacterium *Salmonella typhi*. *Proc Natl Acad Sci USA 92*: 1018-1022.

Majumder R, Sengupta S, Khetawat G, Bhadra R, K, Roychoudhury S, Das J 1996. Physical map of the genome of *Vibrio cholerae* 569B and localization of genetic markers. *J Bacteriol 178*: 1105-1112.

Nandi S, Khetawat G, Sengupta S, Majumder R, Kar S, Bhadra RK, Roychoudhury S, Das J 1997. Rearrangements in the genome of *Vibrio cholerae* strains belonging to different serovars and biovars. *Int J Syst Bacteriol 47*: 858-862.

Popovic T, Bopp CA, Olsvik O, Wachsmuth K 1993. Epidemiologic application of a standardized ribotype scheme for *V. cholerae* O1. *J Clin Microbiol 31*: 2474-2482.

Popovic T, Fields PI, Olsvik O, Wells JG, Evins GM, Cameron DN, Farmer III JJ, Bopp CA, Wachsmuth K, Sack RB, Albert MJ, Nair GB, Shimada T, Feeley JC 1995. Molecular subtyping of toxigenic *Vibrio cholerae* O139 causing epidemic cholera in India and Bangladesh, 1992-1993. *J Infect Dis 171*: 122-127.

Wachsmuth IK, Evins GM, Fields PI, Olsvik O, Popovic T, Bopp CA, Wells JG, Carrillo C, Blake PA 1993. The molecular epidemiology of cholera in Latin America. *J Infect Dis 167*: 621-626.

Yamai S, Okitsu T, Shimada T, Katsube Y 1997. Distribution of serogroups of *Vibrio cholerae* non-O1 non-O139 with specific reference to their ability to produce cholera toxin, and addition of novel serogroups. *J Japan Assoc Infect Dis 71*: 1037-1045.

The Amazonia Variant of *Vibrio cholerae*: Molecular Identification and Study of Virulence Genes

MAS Baptista, JRC Andrade*, ACP Vicente, CA Salles***, A Coelho/+**

Departamento de Genética- I. Biologia, Universidade Federal do Rio de Janeiro, Ilha do Fundão, Cx Postal 68011, 21944-970 Rio de Janeiro, RJ, Brasil *Serviço de Microbiologia e Imunologia, Universidade do Estado do Rio de Janeiro **Departamento de Genética, Instituto Oswaldo Cruz ***Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, Rio de Janeiro, RJ, Brasil

The pathogenic O1 Amazonia variant of Vibrio cholerae *has been shown previously to have a cytotoxin acting on cultured Vero and Y-1 cells, and to lack important virulence factors such as the cholera toxin (Coelho et al. 1995a). This study extends the molecular analysis of the Amazonia strains, detecting the presence of the* toxR *gene, with a very similar sequence to that of the El Tor and classical biotypes. The outer membrane proteins are analyzed, detecting a variation among the group of Amazonia strains, with three different patterns found. As a by-product of this work a polymerase chain reaction fragment was sequenced, reading part of the sequence of the Lon protease of the Amazonia strains. This gene was not previously described in* V. cholerae, *but its sequence is present in the TIGR database specific for this species.*

Key words: *Vibrio cholerae* - Amazonia - *tox*R - outer membrane proteins - protease - Lon

The Amazonia variant of *Vibrio cholerae* was isolated from a group of clinical O1 isolates obtained from the northwest of Brazil in 1991-1992 (Coelho et al. 1995a). The Latin American epidemic was caused by a strain of the El Tor biotype. It spread out from the Pacific coast of Latin America inwards, mainly following the Amazon river basin, reaching the northeast of Brazil and coming down up to the Rio de Janeiro region. A collection of strains from the beginning of the epidemic in Brazil was analyzed by the random amplified polymorphic DNA (RAPD) discriminative technique (Coelho et al. 1995b), and a surprising result was obtained. There was a group of strains with a different fingerprint pattern from the epidemic El Tor strains. All of these strains presented the same pattern, showing that they represented a distinct group.

These 14 strains were analyzed by various techniques. Biochemically they are undistinguishable from other *V. cholerae* strains. All of these strains were Ogawa, in contrast to the majority of strains collected at this time that were Inaba. However other El Tor strains from the same time and area were also Ogawa. The isozyme method (Salles & Momen 1991) was used on these strains, and they were classified into a new group. The same thing happened with ribotypes (Popovic et al. 1993), and they formed a new group.

The Amazonia strains were tested for the presence of the *ctx* gene (Kaper & Levine 1981, Salles et al. 1993, Kaper et al. 1994), encoding the cholera toxin, and the result was negative. The presence of other virulence genes was tested by polymerase chain reaction (PCR), and neither the ST (thermo-stable toxin) (Ogawa et al. 1990, Vicente et al. 1997a) nor the *zot* (zonula occludens) toxin (Baudry et al. 1992) were found. The *tcp*A gene, coding for the colonization pilus (Taylor et al. 1987, Rhine & Taylor 1994, Manning 1997, Vicente et al. 1997b), was not found by PCR or Southern hybridization.

When tested on rabbit ligated ileal loop, the strain did not produce an accumulation of liquid, but did show a destruction of the intestinal ephitelium, a heavy mucus production, with a large number of erythrocytes and epithelial cells embedded in it. In *in vitro* studies on cultured Vero cells, the production of a cytotoxin was detected, leading to morphological alterations of the cells, their detachment from the plastic and death.

In this paper the analysis of virulence genes of the Amazonia strain is extended, mainly with the study of the regulatory gene *tox*R (Peterson & Mekalanos 1988, DiRita & Mekalanos 1991, DiRita et al. 1991). *tox*R is considered a main regulatory gene, responsible for the recognition of en-

This work was supported by a combination of grants from the following institutions: Pronex (CNPq), CAPES, CNPq, Universidade Federal do Rio de Janeiro, Universidade do Estado do Rio de Janeiro, FUJB (Brazil), PAPES (Fiocruz).
+Corresponding author. Fax:+55-21-280-8043. E-mail: coelho@acd.ufrj.br
Received 15 June 1998
Accepted 30 July 1998

vironmental stimuli for expression of a number of genes collectively denominated the ToxR virulence regulon (Skorupski & Taylor 1997, Champion et al. 1997). The major outer membrane protein of *V. cholerae*, OmpU, is directly regulated by the ToxR protein (Miller & Mekalanos 1988, Sperandio et al. 1995, Chakrabarti et al. 1996, Crawford et al. 1998). The outer membrane proteins of the Amazonia strains are also analyzed here.

MATERIALS AND METHODS

Bacterial strains - Fourteen *V. cholerae* Amazonia strains were previously described (Coelho et al. 1995a). A further group of six strains was obtained from Cholera Reference Center (Fiocruz, Brazil). A streptomycin resistant derivative of one of the original strains, 4010 was used for all the experiments described in this paper. Control classical and El Tor strains were O395 and E7946 respectively.

DNA preparation, PCR reaction conditions and product analysis - Bacteria were grown overnight in alkaline peptone water (1 ml) and DNA was extracted (Silhavy et al. 1984). The program used for PCR consisted of 35 cycles, at 94°C for 1 min, 55°C for 1min 30 sec and 72°C for 1min 30 sec. The reactions included 1 µl of each primer (500ng/µl), 100ng of DNA, dNTP's 50 µM each, 5 µl reaction buffer (1.5mM $MgCl_2$ final concentration), 0.5 µl *Taq* polymerase (2.5U) (Pharmacia) and distilled water to a total volume of 50 µl. An MJ Research thermocycler (Watertown, Mass.) was used for the temperature cycling. Primers used for the *tox*R fragment of 560bp were: OL.1: 5' TCGGATTAGGA CACAACTC and OL.2: 5' CTGCGAGGGGAA GTAAGAC. DNA was analyzed on 1.4% agarose gels in TBE 1X, prepared according to Sambrook et al. (1989), and running at 100 Volts for approximately 2 hr 30 min, until the bromophenolblue reached the end of the gel.

Southern transfer and hybridization - Southern transfers to nitrocellulose were done according to Sambrook et al. 1989. The hybridization solution was 50% formamide, 6X SSC, 0.7% SDS. DNA (200ng) was labeled with the *Random Primer* kit (Life Technologies) employing αP32 dCTP.

Plasmid preparations, DNA restriction and ligation - Plasmid preparations employed Qiagen P-100 columns according to instructions of the manufacturers. Restrictions were done as described by the enzyme manufacturers (Life Technologies). A five to one proportion of insert fragment was used in the cloning experiments. Electroporation was done into the *Escherichia coli* strain DH5α.

Outer membrane proteins preparation and protein analysis on SDS-polyacrylamide gels - Outer membrane proteins were prepared from 1 ml of cells. Bacteria were spun down and treated for 10 min with 0.06M Tris HCl (pH 8.0)/ 0.2M sucrose, 0.2mM EDTA and 0.04 mg/ml lysozyme (total volume 500 µl). 10 µl of 1 mg/ml DNAse were added, and then 500 µl of Triton extraction buffer [(2%Triton X-100, 10mM $MgCl_2$, 50mM Tris-HCl (pH 8.0)] were added. Outer membrane fragments were spun down and washed with water for three times. Proteins were resuspended in SDS-PAGE sample buffer, boiled for 5 min, and loaded on 12% SDS-polyacrylamide gels with 5% stacking gels. Electrophoresis was carried out at constant current (35mA), and the gels were stained with 0.25% Coomassie blue, and destained with methanol/acetic acid.

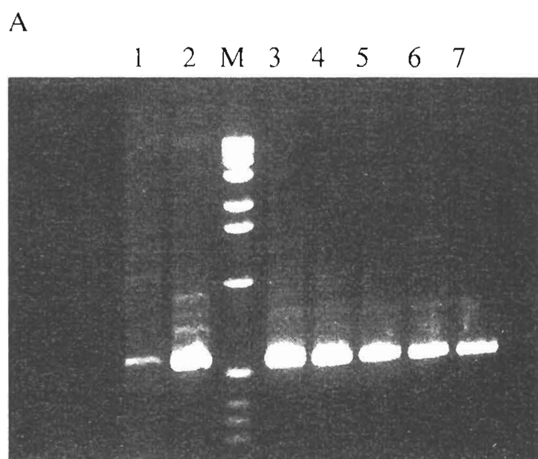
DNA sequencing and analysis - DNA sequencing was done employing the *Thermo-sequenase* kit (Amersham) and αP33 labeled dideoxy nucleotides. Plasmid DNA or PCR amplicons were sequenced. Specific bands on agarose gels were cut, and the DNA purified with the Sephaglass kit (Pharmacia). 300 ng of DNA and 0.5 µl of the ddNTPs were used for sequencing. Standard 60 cm 6% polyacrylamide-bisacrylamide gels were used, with a glycerol tolerant buffer provided with the kit. The gels were fixed and dried and Hyperfilm was exposed for the visualization of the bands.

Databank searches with the sequences were made against the specific *V. cholerae* TIGR database at the Institute for Genomic Research (www.ncbi.nlm.gov/cgi-bin/BLAST/nph-tigrbl) and against the non-redundant combined database through the Blast Search (www.ncbi.nlm.nih.gov/cgi-bin/BLAST/nph-newblast?Jform=0). Individual sequences were retrieved from the Genbank (www2.ncbi.nlm.nih.gov/genbank/query_form.html).

RESULTS

Presence of the toxR gene - Primers OL.1 and OL.2 were used in a test to check whether the *tox*R gene was present. Fig. 1A shows the presence of a strong band with an expected size of 560 bp, for various Amazonia strains. The gel was transferred and hybridized to a probe prepared from a classical strain, by amplification with the same primers. Hybridization was positive, as shown in Fig. 1B.

Cloning of a fragment of the toxR gene - The 560bp PCR fragment of the Amazonia *tox*R was cloned into the pBluescript SK vector, using *Sma*I, and producing the plasmid pMB560R. This fragment was then transferred to a second plasmid, pCVD442 (Donnenberg & Kaper 1991), a suicide plasmid that does not replicate in *V. cholerae*. *Xba*I and *Sal*I were used in the second cloning, and the plasmids obtained were denominated pCVD560R (Fig. 2). A restriction analysis of two such clones (Fig. 3A), and hybridization to the Amazonia *tox*R



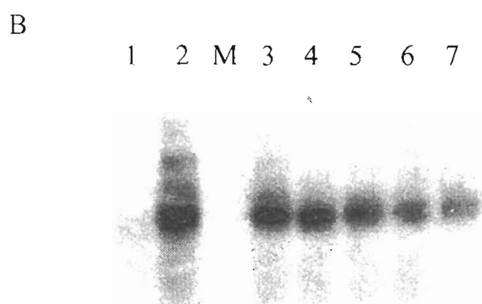


Fig. 1A: *tox*R PCR amplification from Amazonia strains. *tox*R primers were OL.1 and OL.2, which delimit a fragment of 560bp. Lanes 1 and 2 contain products from the *Escherichia coli* DH5α (negative control) and the classical strain O395, respectively. Lanes 3 through 7, PCR products from the Amazonia strains. 3, 4010; 4, 3729; 5, 3506; 6, 3439 and 7, L-34. M indicates the size marker, 1 kb ladder (Life Technologies). Fig. 1B: Southern hybridization of the gel on A with the 560bp *tox*R fragment from the O395 classical strain.

(Fig. 3B) were done, in a structural analysis of the clones.

Partial sequencing of the toxR gene - The *tox*R Amazonia 560bp fragment was sequenced, using both a universal primer for the pMB560R or internal *tox*R primers. A high similarity of the sequence (98.6%) was found to that of the classical biotypes (Fig. 4A). A translation of this sequence shows a 98.4% identity to the classical and El Tor ToxR protein (Fig. 4B). An arginine (R) for threonine 125(T) substitution in particular could cause a difference in the secondary structure of the protein.

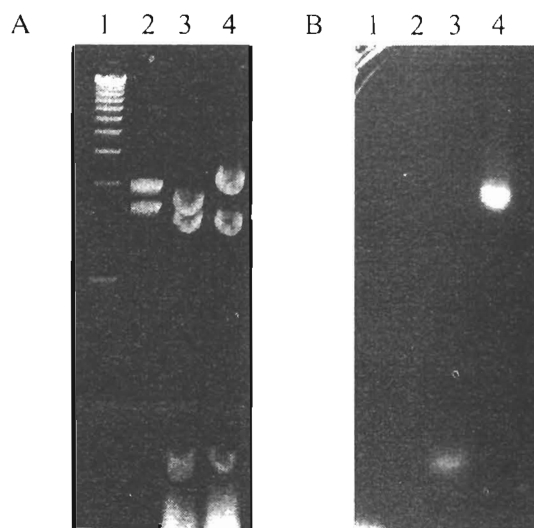


Fig. 3A: *Pst*I restriction analysis of pCVD442 and two different pCVD560R clones. Lane 1, 1kb ladder (Life Technologies); lane 2, pCVD442; lane 3, pCVD560R cl.1; lane 4, pCVD560R cl.2. Fig. 3B: hybridization of the gel on A to a 4010 Amazonia 560bp *tox*R probe.

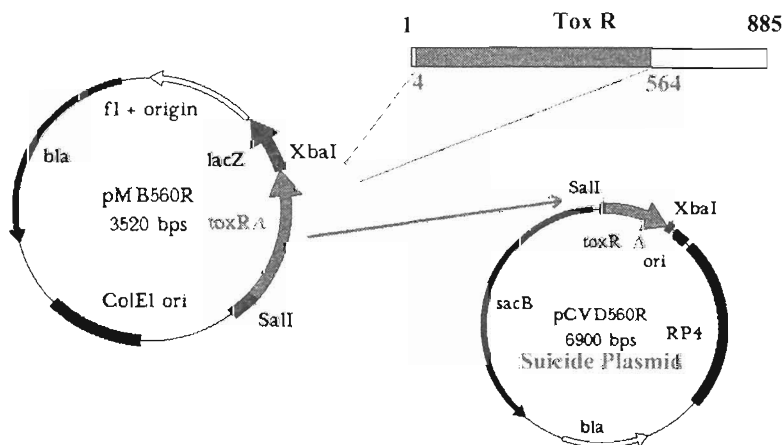


Fig. 2: cloning of the 560bp PCR fragment of the Amazonia strain 4010Sm^r into the *Sma*I site of pBluescript SK, producing plasmid pMB560R and further cloning, with *Xba*I and *Sal*I, into pCVD442, to yield plasmids pCVD560R.

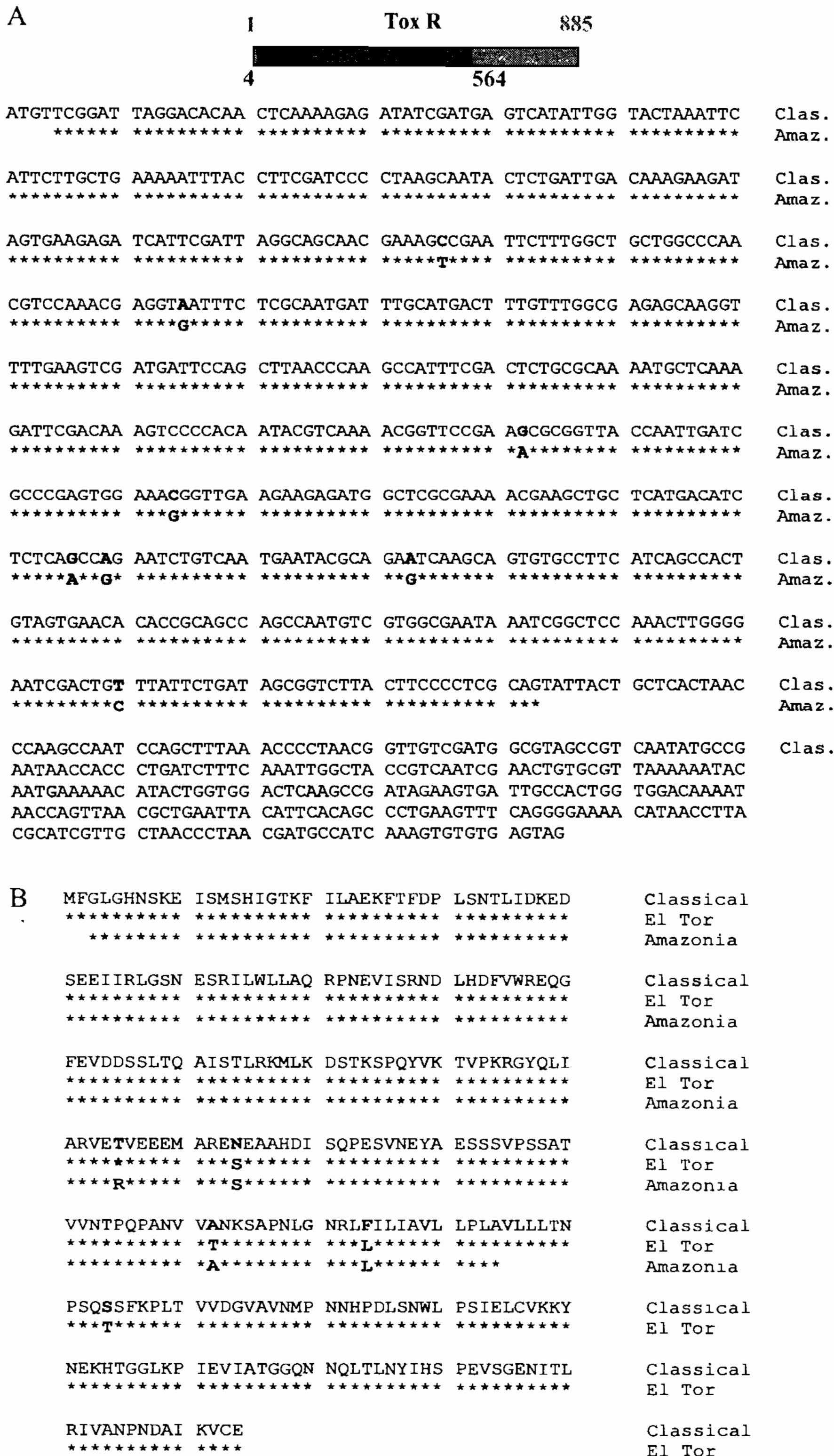


Fig. 4A: DNA sequence of the *tox*R fragment from the Amazonia strain 4010 compared to the sequence of the classical 569B strain. Numbers correspond to the number of nucleotides. * are used to mark the same nucleotide as in the previous line. Fig. 4B: aminoacid comparison between a translation of the DNA sequence of the Amazonia strain 4010, compared to the sequence of the ToxR protein of classical strain 569B and El Tor strain E7946. * are used to mark the same aminoacid as in the previous line.

The major outer membrane proteins of Amazonia strains - The outer membrane proteins of the Amazonia strains were analyzed on polyacrylamide gels (Fig. 5). Three types of patterns were obtained. Some strains present one major protein with a size of 38kDa, as the 38kDa OmpU protein of the El Tor and classical strains. Other strains present one smaller protein of 35kDa. A third group presents both of these proteins. Different media were used for growth, to check whether this major proteins would vary according to growth conditions (data not shown), but it seems that the pattern is strain determined.

1 2 3 4 5 6 7 8 9 10

Fig. 5: SDS- polyacrylamide gel electrophoresis of outer membrane proteins. Lane 1: classical strain O395; 2-6, Amazonia variant strains (2, L-34; 3, 3218; 4, 3439; 5, 4010; 6, 3506); 7, El Tor (607); 8 and 9, Amazonia variant strains (8, 3501; 9, 4132); 10, protein molecular weight markers.

Presence of the lon gene - In the process of looking for virulence related genes, we obtained a spurious 800bp PCR amplicon that was partially sequenced. This sequence corresponds to the Lon protease of various bacteria. This is one of the main proteases, responsible for the degradation of misfolded proteins (Gottesman 1996). In Fig. 6 we compare the sequence of the Amazonia strain putative Lon protein to a sequence found in the TIGR databank for *V. cholerae* and to the other bacteria and archae.

DISCUSSION

The pathogenic O1 Amazonia *Vibrio cholerae* may be considered a new human pathogen. It was discovered with the use of discriminative molecular techniques (Coelho et al. 1995b, 1997). This strain belongs to the same species as the epidemic strains but seems to cause diarrhea by a different mechanism. The lack of the cholera toxin, and the presence of a cytotoxin leading to intestinal tissue damage in rabbits strongly suggest a different route for the disease.

The *tox*R gene was found in the Amazonia strains, and its sequence is highly homologous to that of the El Tor and classical biotypes. A *tox*R homologous gene has been described in *Vibrio parahaemolyticus*, *Vibrio fischeri* and a *Photobacterium* sp., with much lower DNA homology values, all below 70%.

The presence of the regulatory gene *tox*R, in the absence of various genes that it normally regulates, leads to the question of its further role. One possibility is the regulation of the outer membrane

```
E. coli             EQQIPFSASLTFEQSYSEVDGDSASMAELCALISALADVPVNQSIAITGSVDQFGRAQPV
                       ::     ::::::  :::::::::: ::  ::    :  : ::::::   ::  :
V. cholerae         TARIPLTTTITFEQSYGGVDGDSASMAEFCAIVSAFSKQPNRQDIAITGSMNHFGESQAI
                           ::::::::::::::::::::  ::::::::::::::::::::
V.chol. Amazonia           TTITFEQSYGGVDGDSASMA?LCAIVSAFSKQPNRQDIAITGRLR?K
                               ::::::  :::::: :  :  :::    :  : :::::
Hae. inf.           PSQLPFSASLVFEQSYGEIDGDSASLAIFCVLVSALADLPLPQHIAITGSIDQFGLVHSV
                               :   :    :::::  :     ::       :  : :::    : :  :
Arch.fulgidus        DISNMDVHIQFVGTYEGVEGDSASISIATAVISAIEGIPVDQSVAMTGSLSVKGEVLPV
                     :::: : ::::   : :::::::: : ::::::: : :::::::: :::::  : ::::
Meth.thermo.        TDISNYDIHIQFLQAYDGVEGDSASVSVATAVISALEEIPVDQSVALTGSLSIRGDVLPV
```

Fig. 6: aminoacid sequence comparison of the Amazonia putative Lon protease and the corresponding protein in other bacteria and archae. Bacteria: *Escherichia coli*, *Vibrio cholerae*, *Haemophilus influenzae*. Archae: *Archaeoglobus fulgidus* and *Methanobacterium thermoauxotrophicum*.

proteins. It is known that *tox*R directly regulates the OmpU protein, which has been proposed as an adhesin with a role in virulence (Sperandio et al. 1995, 1996). The Amazonia strain, as in the case of El Tor and classical strains, presents major outer membrane proteins, which could have functional homologies to OmpU. Further developments in this work will include studies on the presence of other regulatory genes such as genes of the tcp cluster, like the *tcp*P and *tcp*H (Carroll et al. 1997, Manning 1997, Häse & Mekalanos 1998) and other putative virulence genes (Karaolis et al. 1998, Tacket et al. 1998). It will also aim at the construction of *tox*R mutants to evaluate the role of this gene in the pathogenicity of the Amazonia strain, including an analysis of its adhesion to cells, cytotoxicity, and behavior in rabbit ileal loops. These comparative studies may shed light not only on the virulence mechanisms involved in the Amazonia strain but also on *V. cholerae* itself.

ACKNOWLEDGMENTS

To the Cholera Reference Center (Fiocruz) that provided a second group of six strains that we typed as belonging to the Amazonia group

REFERENCES

Baudry B, Fasano A, Ketley J, Kaper JB 1992. Cloning of a gene (*zot*) encoding a new toxin produced by *Vibrio cholerae*. *Infect Immun 60*: 428-434.

Carroll PA, Tashima KT, Rogers MB, DiRita VJ, Calderwood SB 1997. Phase variation in tcpH modulates expression of the ToxR regulon in *Vibrio cholerae. Mol Microbiol 25*: 1009-1011.

Chakrabarti SR, Chaudhuri K, Sem K, Das J 1996. Porins of *Vibrio cholerae*: purification and characterization of OmpU. *J Bacteriol 178*: 524-530.

Champion GA, Neely MN, Brennan MA, DiRita VJ 1997. A branch in the ToxR regulatory cascade of *Vibrio cholerae* revealed by characterization of *tox*T mutant strains. *Mol Microbiol 23*: 323-331.

Coelho A, Andrade JRC, Baptista MAS, Vicente ACP, Ferraz LR, Salles CA 1997. Genomic fingerprints of *Vibrio cholerae* using arbitrary primer PCR are useful epidemiological tools, p. 203-211. In GT Keusch, M Kawakami (eds), *Cytokines, Cholera and the Gut*, IOS Press, Netherlands.

Coelho A, Andrade JRC, Vicente ACP, Salles CA 1995a. A new variant of *Vibrio cholerae* O1 from clinical isolates in Amazonia. *J Clin Microbiol 33*: 114-118.

Coelho A, Vicente ACP, Baptista MAS, Momen H, dos Santos FRW, Salles CA 1995b. The distinction of pathogenic *Vibrio cholerae* groups using arbitrarily primed PCR fingerprints. *Res Microbiol 146*: 671-683.

Crawford JA, Kaper JB, DiRita VJ 1998. Analysis of ToxR-dependent transcription activation of *ompU*, the gene encoding a major envelope protein in *Vibrio cholerae*. *Mol Microbiol* In press.

DiRita VJ, Mekalanos JJ 1991. Periplasmic interaction between two membrane regulatory proteins, ToxR and ToxS, results in signal transduction and transcriptional activation. *Cell 64*: 29-37.

DiRita V, Parsot C, Jander G, Mekalanos JJ 1991. Regulatory cascade controls virulence in *Vibrio cholerae*. *Proc Natl Acad Sci USA 88*: 5403-5407.

Donnenberg MS, Kaper JB 1991. Construction of an *eae* deletion mutant of enteropathogenic *Escherichia coli* by using a positive- selection suicide vector. *Infect Immun 59*: 4310-4317.

Gottesman S 1996. Proteases and their targets in *Escherichia coli*. *Ann Rev Genet 30*: 465-506.

Häse CC, Mekalanos JJ 1998. TcpP is a positive regulator of virulence gene expression in *Vibrio cholerae*. *Proc Natl Acad Sci USA 95*: 730-734.

Kaper JB, Fasano A, Trucksis M 1994. Toxins of *Vibrio cholerae*, p. 145-176. In IK Wachsmuth, PA Blake, Ø Olsvik (eds), *Vibrio cholerae and cholera. molecular to global perspectives*, ASM Press, Washington.

Kaper JB, Levine MM 1981. Cloned cholera enterotoxin genes in study and prevention of cholera. *Lancet ii*: 1162-1163.

Karaolis DK, Johnson JA, Bailey CC, Boedeker EC, Kaper JB, Reeves PR 1998. A *Vibrio cholerae* pathogenicity island associated with epidemic and pandemic strains. *Proc Natl Acad Sci USA 95*: 3134-3139.

Manning PA 1997. The *tcp* gene cluster of *Vibrio cholerae Gene 192*: 63-70.

Miller VL, Mekalanos JJ 1988. A novel suicide vector and its use in construction on insertion mutations: osmoregulation of outer membrane proteins and virulence determinants in *Vibrio cholerae* requires *tox*R *J Bacteriol 170*: 2575-2583.

Ogawa A, Kato J-I, Watanabe H, Nair BG, Takeda T 1990. Cloning and nucleotide sequence of a heat-stable enterotoxin gene from *Vibrio cholerae* non-O1 isolated from a patient with traveler's diarrhea. *Infect Immun 58*: 3325-3329.

Peterson KM, Mekalanos JJ 1988. Characterization of the *Vibrio cholerae* ToxR regulon: identification of novel genes involved in intestinal colonization. *Infect Immun 56*: 2822-2829.

Popovic T, Bopp CA, Olsvik Ø, Wachsmuth K 1993. Epidemiologic application of a standardized ribotype scheme for *Vibrio cholerae*. *J Clin Microbiol 31*: 2474-2482.

Rhine JA, Taylor RK 1994. TcpA pilin sequences and colonization requirements for O1 and O139 *Vibrio cholerae*. *Mol Microbiol 13*: 1013-1020.

Salles CA, Momen H 1991. Identification of *Vibrio cholerae* by enzyme electrophoresis. *Trans R Soc Trop Med Hyg 85*: 544-547.

Salles CA, Momen H, Vicente ACP, Coelho A 1993. *Vibrio cholerae* in South America: polymerase chain reaction and zymovar analysis. *Trans R Soc Trop Med Hyg 87*: 272.

Sambrook J, Fritsch EF, Maniatis T 1989. *Molecular Cloning, a Laboratory Manual*. 2nd ed. Cold Spring Harbor Lab Press, New York.

Silhavy TJ, Berman M L, Enquist LW 1984. DNA extraction from bacterial cells, p. 137-139. In *Experiments with Gene Fusions*, Cold Spring Harbor Lab Press, New York, 303 pp.

Skorupski, K, Taylor RK 1997. Control of the ToxR virulence regulon in *Vibrio cholerae* by environmental stimuli. *Mol Microbiol 25*: 1003-1009.

Sperandio V, Bailey C, Girón JA, DiRita VJ, Silveira WD, Vettore AL, Kaper JB 1996. Cloning and characterization of the gene encoding the OmpU outer membrane protein of *Vibrio cholerae. Infect Immun 64*: 5406-5409.

Sperandio V, Girón JA, Silveira WD, Kaper JB 1995. The OmpU outer membrane protein, a potential adherence factor of *Vibrio cholerae. Infect Immun 63*: 4433-4438.

Tacket CO, Taylor RK, Losonsky G, Lim Y, Nataro JP, Kaper JB, Levine MM 1998. Investigation of the roles of toxin- coregulated pili and manose-sensitive hemagglutinin pili in the pathogenesis of *Vibrio cholerae* O139 infection. *Infect Immun 66*: 692-695.

Taylor RK, Miller VL, Furlong DB, Mekalanos JJ 1987. Use of *pho*A gene fusions to identify a pilus colonization factor coordinately regulated with cholera toxin. *Proc Natl Acad Sci USA 84*: 2833-2837.

Vicente ACP, Coelho A, Salles CA 1997a. Detection of *Vibrio cholerae* and *Vibrio mimicus* heat- stable toxin gene sequence by PCR. *J Med Microbiol 47*: 398-402.

Vicente ACP, Novaes RC, Coelho A, Salles CA 1997b. Molecular characterization of TCP cluster in non-O1, non-O139 *Vibrio cholerae*, p. 243-248. In GT Keusch, M Kawakami (eds), *Cytokines, Cholera and the Gut*, IOS Press, Netherlands.

Molecular Epidemiology and Emergence of Rift Valley Fever

AA Sall/+, PMA Zanotto*, P Vialat, OK Sène, M Bouloy****

Institut Pasteur de Dakar, BP 220, Dakar, Senegal *Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, Cidade Universitária, 05508-900 São Paulo, SP, Brasil
**Institut Pasteur, 25, Rue du docteur Roux, 75724 Paris cedex 15, France

Rift Valley fever (RVF) is a mosquito-borne viral disease which manifested itself during recent epidemics and revealed its significant potential of emergence. Studies on molecular epidemiology undertaken to better understand the factors leading to RVF emergence, have confirmed the mode of circulation of the virus and highlithted probable risks and obstacles for prevention and control. As for several other viral agents, molecular epidemiology is becoming a useful tool in the study of the emergence of RVF as a serious infectious disease.

Key words: arboviruses - Rift Valley fever - molecular epidemiology - emergence - phylogeny

Rift Valley fever (RVF) is an arboviral disease transmitted by mosquitoes in Africa. RVF affects primarily ruminants causing high mortality in offspring and abortions in pregnants females and occasionally humans, whose infection leads to a clinical picture which ranges from a mild febrile case to hemorragic fever with complications such as hepatitis, encephalitis and retinitis (Laughlin et al. 1979). In 1977, a severe outbreak of RVF occurred in human and livestock populations of Egypt (Meegan 1981). Although RVF was known for more than 40 years at that time, the extensive morbidity and mortality observed in humans appeared as a novelty in the history of this disease, therefore, emphasizing RVF as a serious emerging threat for humans and animals health. Futher large scale epidemics in Mauritania (Digoutte & Peters 1989), Madagascar (Morvan et al. 1991, 1992a, b), Egypt (Arthur et al. 1993) and very recently in eastern Africa (Anonymous 1998) confirmed the major impact of RVF on public health through its continuing emergence. Thus, RVF constitutes an excellent model to overview factors involved in arboviruses emergence because most of the concepts relative to emerging diseases may be illustrated along its natural history.

Control of RVF implies the better identification of factors involved in its emergence and its maintenance in nature. It is also necessary to understand the rules and modalities of circulation and evolution of RVF virus (RVFV) in Africa. These latter objectives have been addressed by studying the variability among RVFV isolates by serological (Besselar et al. 1991) and molecular methods (Battles & Dalrymple 1988, Sall et al. 1997a, b). This paper aims the discussion of some of the aspects and contributions of molecular epidemiology towards the elucidation of RVF emergence.

BACKGROUND

Discovery of RVFV and recent major epidemic/epizootics

RVFV was first isolated in 1930 near lake Naivasha in Kenya by Daubney et al. (1931). Since then, the virus has been shown to be widespread in subsaharian Africa and in Egypt (Meegan & Bailey 1989). Major epidemic/epizootics occurred in Egypt in 1977 (200,000 humans infections and 600 deaths) and in 1993, Mauritania in 1987 (200 human deaths), Madagascar in 1991 and in eastern Africa (89,000 infections and more than 500 deaths reported so far) with the last recent outbreak in 1997-1998 in Kenya, Tanzania, Somalia.

The etiological agent of RVF

RVFV is a member of Bunyaviridae family, *Phlebovirus* genus (Murphy et al. 1995). Its genome consists in three negative single stranded RNA segments referred as L, M and S respectively for large, medium and small. The L segment codes for the L protein which is the viral polymerase. The M segment codes for glycoproteins G1 and G2 and two others proteins of 78 and 14 K. The S segment codes for the nucleoprotein N and the non structural NSs protein using an ambisense strategy (Bouloy 1991, Elliott et al. 1991, Giorgi 1996, Schmaljohn 1996).

+Corresponding author. Fax: +221.839.92.10. E-mail: sall@pasteur.sn
Received 15 June 1998
Accepted 30 July 1998

Epidemiology of RVF

The epidemiology of RVF consists in both epizootic and interepizootic cycles (Meegan & Bailey 1989). Epizootics of RVF in Africa occured often when unusually heavy rainfall were observed. During an epizootic, virus circulates among infected arthropod vectors and mammalian hosts, particularly cattle and sheep, which represent the most significant livestock amplifiers of RVFV. The inter-epizootic survival of RVFV is believed to depend on transovarial transmission of virus in floodwater *Aedes* mosquitoes (Linthicum et al. 1985). Virus can persist in mosquitoes eggs until the next period of heavy rainfall when they hatch and yield RVFV infected mosquitoes. Depending on factors such as availability of sufficient numbers of competent mosquito vectors, presence of susceptible vertebrates, appropriate environmental conditions, infected mosquitoes have the potential to infect a relatively small number of vertebrate hosts or to initiate a widespread RVF epizootic.

Control and prevention of RVF

Vaccines have been the principal mean used to control RVF. Two types of vaccines have been described for use against RVF: inactivated and live-attenuated.

Formalin-inactivated RVF vaccines have been used to immunize animals, laboratory workers, veterinarians and other people at high risk of exposure to RVFV. The cost of the vaccine, the requirement for multiple inoculations and the time interval required to mount a protective immune response, all limit its use for veterinary purposes.

Two live attenuated vaccines, the Smithburn vaccine, also referred as Smithburn neurotropic strain or SNS (Smithburn et al. 1949), and MP12 (Caplen et al. 1985) have been developped. The Smithburn strain is the only widely available veterinary vaccine but has serious limitations in practical use, because it has been proven to be teratogenic, cause abortions and encephalitis in young lambs.

Possessing attenuation markers in all three segments, MP12 has a very low probability of reversion (Saluzzo & Smith 1990, Vialat et al. 1997) and has been inoculated into more than 100 people and shown to be safe and immunogenic (Peters 1997). MP12 was also promising in laboratory trials in domestic animals (Morrill et al. 1987, Morril & Mc Clain 1996), but vaccination of pregnant ewes revealed that the virus caused teratogenic effect if inoculated during the first trimester of pregnancy (Erasmus and Bishop, pers. commun.). Another attenuated virus, clone 13, a naturally attenuated strain, is very promising regarding the preliminary results obtained in terms of immunogenicity and safety (Muller et al. 1995).

MOLECULAR EPIDEMIOLOGY OF RVFV

Investigation on the variation among RVFV isolates using serological tests based on the antigenicity of structural proteins (Saluzzo et al. 1989a,b, Besselar et al. 1991) or genetic methods such as T1-oligonucleotide fingerprints (Peters & Linticum 1994) and, more recently sequencing, (Battles & Dalrymple 1988) indicated only minor variations among RVFV natural isolates. To further analyze the genetic diversity of RVFV (Sall et al. 1997b), we selected a panel of 18 strains (Table I) isolated over some 50 years from various hosts and geographical origins and we sequenced directly their NSs coding region on the S segment after a step of reverse transcription-polymerase chain reaction amplification (RT/PCR). A 50% majority rule consensus tree derived from the sequences analyzed are presented in Fig. 1. The NSs coding region sequences clustered in three major lineages supported by high bootstrapping values and by using different phylogenetic inference procedures (e.g., maximum likelihood, parsimony and distance methods) and correlated with the geographic origin of the isolates and are referred as West Africa, East-Central Africa and Egypt. While the West Africa group was homogenous with strains from Mauritania, Senegal, Guinea and Burkina Faso, the East-Central Africa and Egypt ones appeared to be heterogenous.

As expected, the Egyptian group contains strains isolated in 1977 and 1993 epidemics, which appear in the phylogeny as sister groups suggesting that either the virus remained endemic between the two outbreaks or have been reintroduced in 1993 from the same source (probably Sudan) as in 1977. To explain the reemergence of RVF in Egypt after years of silence despite intensive surveillance, Peters (1997) proposed that the virus was reintroduced through an incompletely inactivated RVF veterinary vaccine. Furthermore, Ar MAD 79, which is the first isolation of RVFV in Madagascar, clustered in Egypt group and then confirmed data obtained by Morvan et al. (1991) who analyzed the antigenic properties of the N protein.

Secondly, the East-Central African group clustered isolates from Uganda, Central African Republic, Mauritania and Senegal. The presence of An MAD 91 in that group suggested that this latter strain was probably introduced in Madagascar from the eastern coast of Africa. This latter assumption also implies that there was at least two introductions of the virus in Madagascar but also several lineages coexist in East Africa. Moreover, in

TABLE I

Characteristics of the Rift Valley fever virus isolates analyzed by sequencing

Code	Strain	Year of isolation	Origin	Source
SNS	Smithburn	1944	Uganda	Entebbe strain
Ar UG 55	Lunyo	1955	Uganda	Mosquito
Ar RCA 69	Ar B 1976	1969	CAR	Mosquito
H EGY 77	ZH 548	1977	Egypt	Human
MP 12[a]	MP12	1985	Egypt	ZH548 strain
Ar MAD 79	Ar Mg 811	1979	Madagascar	Mosquito
Ar SEN 84	Ar D 38661	1984	Senegal	Mosquito
An GUI 84	An K 6087	1984	Guinea	Bat
Ar BUF 84	Ar D 38457	1984	Burkina Faso	Mosquito
H1 MAU 87	H D 47502	1987	Mauritania	Human
H2 MAU 87	H D 47311	1987	Mauritania	Human
H3 MAU 87	H D 47408	1987	Mauritania	Human
H4 MAU 87	H D 48255	1987	Mauritania	Human
An MAD 91	An Mg 990	1991	Madagascar	Bovine
Ar SEN 93	Ar D 104769	1993	Senegal	Mosquito
An SEN 93	An D 106417	1993	Senegal	Zebu
B EGY 93	B EGY 93	1993	Egypt	Buffalo
H EGY 93	H EGY 93	1993	Egypt	Human

a: laboratory-attenuated strain derived from a wild strain; SNS: Smithburn neurotropic strain; H: human; Ar: arthropode; An: animal; B: buffalo.

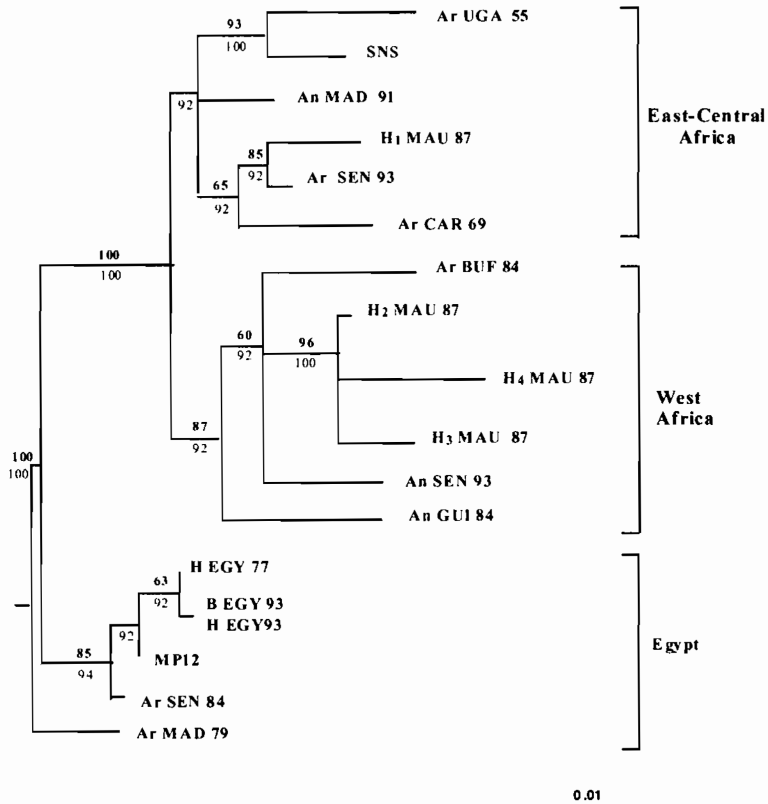


Fig. 1: phylogenetic tree for the NSs gene of several Rift Valley fever virus isolates. Values above branches indicate the level (%) of bootstrap support using maximum parsimony after 500 iterations. Values below branches indicate the number of times a given node was observed on a majority rule consensus of 50 trees with equivalent likelihood (LnL). Branch lengths are shown proportional to the number of substitutions per 100 residues. The rooting shown here was determined by the inclusion of the SSF NSs sequence.

Uganda, despite 11 years separating the isolations of Ar UGA 55 and Entebbe strain, the parental strain of Smithburn vaccinal strain, RVFV did not show much genetic diversity, suggesting a maintenance mechanism through an endemic/enzootic cycle, possibly involving comparatively little viral activity, since increased genetic diversity for a given mutation rate entails an increase on the effective viral population size. Surprisingly, H1 MAU 87 and Ar SEN 93 belonged to the East-Central Africa group.

The West Africa group appeared to be homogenous and suggested circulation of similar variants in Senegal, Mauritania, Guinea and Burkina Faso. It is noteworthy that H2, 3, and 4 MAU 87 clustered near each other and were isolated from fatal cases whereas H1 MAU 87 which was isolated from a febrile case clustered together with Ar SEN 93 unexpectedly in the East-Central Africa lineage. Moreover, one may deduced, from the strains distribution on Fig. 1, that there are two areas of circulation of RVFV in Senegal: (i) the Northern Sahelian zone where Ar SEN 93 and H1 MAU 87 were isolated and, (ii) the Sudano-Guinean zone where An SEN 93 was isolated.

Groupings of Ar SEN 84 with Egyptian strains on one hand and H1MAU 87 and Ar SEN 93 with eastern and central African strains on the other hand, were quite unexpected and led us to hypothesize genetic exchange through reassortment to explain these puzzling clusterings. In order to check this hypothesis further sequencing and phylogenetic analysis were undertaken both on L and M segments. Although, this hypothesis is still under investigation, one may obviously speculate by anticipation that such a mechanism *in natura* would have important implications on epidemiology and emergence of RVF in Africa (see below).

TABLE II

Summary of Rift Valley fever (RVF) emergence factors described and analyzed by Wilson (1994)

Factors of emergence	Examples relative to RVF
Economic development/ Land use	Dams and irrigation, pasturage improvement
Human demography and behavior	Living with domestic ungulates, slaughter of sick animal, Vaccination of healthy animals
International travel and commerce	Domestic ungulates export, human travel and migration
Biological adaptation and changc	Increased viral virulence, improved vector competence, greater animal susceptibility
Climate events	Excessive rainfall

EMERGENCE OF RVFV

Various factors contributing to the emergence of infectious diseases were classified by Lederberg et al. (1992) and analyzed from the point of view of RVF by Wilson (1994) and summarized in Table II. Emergence of RVF was also discussed by Peters (1997) with special reference to Madagascar, distant spread of the virus to Egypt and historical speculations. These two papers emphasized the multifactorial aspect of RVF emergence and the central role of water and ecological change as factors triggering epidemics. Water is usually involved either through dams or irrigation for the sake of agriculture development, as illustrated by Egypt in 1977 and Mauritania in 1987 or, under excessive rainfall and floodings as observed during the 1997-98 outbreak in eastern Africa. Concerning the impact of ecological changes as deforestation and agricultural pratices change, the outbreak in Madagascar in 1991 has been shown to be a very instructive example (Peters & Linthicum 1994, Peters 1997).

Although these two key factors were clearly identified and characterized, data derived from molecular epidemiology are needed for a comprehensive view of RVF and its emergence process. Our work, although still preliminary allowed to illustrate the contribution of molecular epidemiology for, (i) the understanding of two modes of circulation of viral strains and (ii) delineation of genetic aspects of the virus, which may turn out to become potential obstacles for the prevention and control of the disease.

Modes of circulation of RVFV

Regarding the molecular epidemiology data about RVFV, two modes of circulation may be illustrated (Fig. 2): (i) distant spread from one region to another and (ii) local circulation in an enzootic/endemic area.

Distant spread was illustrated by introduction of RVFV in Egypt (1977) and Madagascar (1979 and 1991) probably from eastern or central Africa (see molecular epidemiology). It is interesting to emphasize that in both cases, possibly an antigenically and phylogenetically "new" virus was introduced in an area exempt of RVFV, raising the issue about the role of herd immunity for both, humans and animals populations, as a factor of emergence of the virus.

Concerning local circulation in an enzootic/endemic area, Senegal is an instructive example

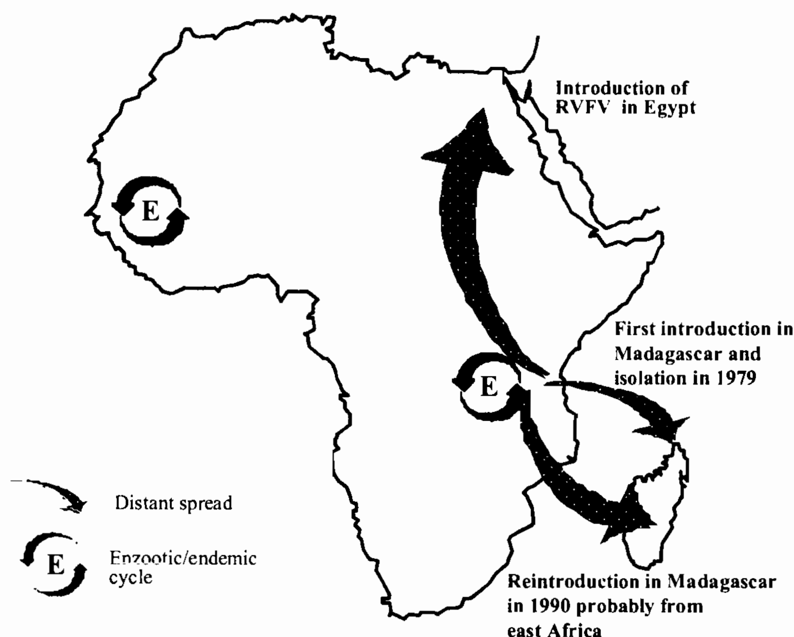


Fig. 2: possible modes of circulation of Rift Valley fever.

because it showed reemergence of the virus from a pool of existing enzootic/endemic strains under a similar process although the ecological context of transmission is different between the north and the south of that country. As far as RVF is concerned, Senegal can be divided in two areas (Sall et al. 1997b) from which the virus have emerged as demonstrated by isolations in 1993 (Zeller et al. 1997): (i) the Sahelian zone, where southern Mauritanian and northern Senegalese strains are circulating and, (ii) the Sudano-Guinean zone where southern Senegalese strains are in contact with those from bordering countries.

Prevention and control

In the field of prevention and control of RVFV, molecular epidemiology studies highlighted a potential major obstacle to the use of live attenuated vaccines. Indeed, the possibility of the existence of reassortment in nature raised by the unexpected groupings (Ar SEN 84, H1 MAU 87 and Ar SEN 93) would emphasize the risk of generating uncontrolled chimeric viruses.

CONCLUSION

Although molecular epidemiology has been shown to be informative for a better understanding on different facets of RVFV emergence, many questions such as those relative to the sylvatic cycle of the virus for instance remain unanswered. Meanwhile, surveillance of RVF and awareness should be improved and reinforced since it is so far the only conceivable way to prevent RVFV emergence with its toll of deaths, sickness and economic loss.

ACKNOWLEDGEMENTS

To C Mathiot and J Thonnon for critically reading the manuscript.

REFERENCES

Anonymous 1998. An outbreak of Rift Valley fever, Eastern Africa, 1997-1998. *WER 73*: 105-112.

Arthur R, El-Sharkawy MS, Cope SE, Botros BA, Oun S, Morrill JC, Shope RE, Hibbs RG, Darwish MA, Imam IZE 1993. Recurrence of Rift Valley fever in Egypt. *Lancet 342*: 1149-50.

Battles JK, Dalrymple JM 1988. Genetic variation among geographic isolates of rift valley fever virus. *Am J Trop Med Hyg 39*: 617-631.

Besselar TG, Blackburn NK, Meenehan GM 1991. Antigenic analysis of Rift Valley fever isolates: monoclonal antibodies distinguish between wild-type and neurotropic virus strains. *Res Virol 142:* 469-474.

Bouloy M 1991. Bunyaviridae: Genome organization and replication strategies. *Adv Vir Res 40*: 235-266.

Caplen H, Peters CJ, Bishop DHL 1985. Mutagen-directed attenuation of Rift Valley fever virus as a method for vaccine development. *J Gen Virol 66*: 2271-2277.

Daubney R, Hudson JR, Garnham PC 1931. Enzootic hepatitis or Rift Valley fever. An undescribed virus disease of sheep, cattle and man from East Africa. *J Pathol Bacteriol 34*: 545-79.

Digoutte JP, Peters CJ 1989. General aspects of the 1987 Rift Valley fever epidemic in Mauritania. *Res Virol 140*: 27-30.

Elliott RM, Schmaljohn CS, Collett MS 1991. Bunyaviridae genome structure and gene expression. *Curr Top Micro Immunol*, Springer- Verlag, Berlin: 91-142.

Giorgi C 1996. Molecular biology of phleboviruses, p. 105-128. In RM Eliott, *The Bunyaviridae*, Plenum Press, New York.

Laughlin LW, Meegan JM, Strausbaugh LW, Morens DM, Watten RH 1979. Epidemic Rift Valley fever in Egypt: observations of the spectrum of human illness. *Trans R Soc Trop Med Hyg 73*: 630-633.

Lederberg J, Shope RE, Oaks SC 1992. *Emerging Infections, Microbial Threats to Health in the United States*, National Academy Press, Washington, DC.

Linthicum KJ, Davies FG, Kairo A 1985. Rift Valley fever virus (family Bunyaviridae, genus *Phlebovirus*). Isolations from Diptera collected during an inter-epizootic period in Kenya. *J Hyg Camb 95*: 197-209.

Meegan JM 1981. Rift Valley fever in Egypt. An overview of epizootics in 1977 and 1978. *Contrib Epidemiol Biostat 3*: 100-113.

Meegan JM, Bailey CJ 1989. Rift Valley fever, p. 51-76. In TP Monath, *The Arboviruses: Epidemiology and Ecology IV*, CRC Press Inc., Boca Raton, Fl.

Morvan J, Saluzzo JF, Fontenille D, Rollin PE, Coulanges P 1991. Rift Valley fever in the east coast of Madagascar. *Res Virol 142*: 475-482.

Morvan J, Rollin PE, Laventure S, Rakotoarivony I, Roux J 1992a. Rift Valley fever epizootic in the central highlands of Madagascar. *Res Virol 143*: 407-415.

Morvan J, Rollin PE, Roux J 1992b. La fièvre de la Vallée du Rift à Madagascar en 1991. Enquête séro-epidémiologique chez les bovins. *Rev Ele Med Pay Trop 45*: 121-127.

Morrill JC, Mc Clain DJ 1996. Epidemiology and pathogenesis of Rift Valley fever and other phleboviruses. p. 281-294 In RM Elliott, *The bunyaviridae*, Plenum Press, New York.

Morrill JC, Jennings GB, Caplen H, Turell MJ, Johnson AJ, Peters CJ 1987. Pathogenicity and immunogenicity of a mutagen-attenuated Rift Valley fever virus immunogen in pregnant ewes. *Am J Vet Res 48*: 1042-1047.

Muller R, Saluzzo JF, Lopez N, Dreier T, Turell M, Smith J, Bouloy M 1995. Characterization of clone 13, a naturally attenuated avirulent isolate of Rift Valley fever virus, which is altered in the small segment. *Am J Trop Med Hyg 53*: 405-411.

Murphy FA, Fauquet CM, Bishop DHL, Ghabrial SA, Jarvis AW, Martelli GP, Mayo MA, Summers MD 1995. Family Bunyaviridae, p. 300-315. In *Virus Taxonomy. Classification and Nomenclature of Viruses*. Sixth report of the international committee on taxonomy of viruses. Springer-Verlag, Wien, New York.

Peters CJ 1997. Emergence of Rift Valley fever, p. 253-264. In JF Saluzzo, B Dodet (eds), *Factors in the Emergence of Arbovirus Diseases*, Elsevier, Paris.

Peters CJ, Linthicum KJ 1994. Rift Valley fever, p. 125-138. In Beran GW, Steele JH (eds), *Hanbbook of Zoonoses, Section B: Viral*, 2nd ed., CRC Press, Boca Raton, Fl.

Sall AA, Zeller H, Zanotto PM de A, Billecocq A, Vialat P, Préhaud C, Yadani F, Bouloy M 1997a. Rift Valley Fever virus: the NSs protein as a target to evaluate genetic variability and its roles in pathogenesis, p. 265-271. In JF Saluzzo, B Dodet (eds), *Factors of Emergence of Arboviruses*, Elsevier, Paris.

Sall AA, Zeller H, Thiongane Y, Digoutte JP, Bouloy M 1997b. Genetic variability of the NSs protein among Rift Valley Fever Virus isolates. *J Gen Virol 78*: 2853-2858.

Saluzzo JF, Smith JF 1990. Use of reassortants viruses to map attenuating and temperature sensitive mutations of the Rift Valley fever virus MP12 vaccine. *Vaccine 8*: 369-75.

Saluzzo JF, Anderson GW, Smith JF, Fontenille D, Coulanges P 1989a. Biological and antigenic relationship between Rift Valley fever virus strains isolated in Egypt and Madagascar. *Trans R Soc Trop Med Hyg 83*: 701.

Saluzzo JF, Anderson GW, Hodgson LA, Digoutte JP, Smith JF 1989b. Antigenic and biological properties of Rift Valley fever isolated during the 1987 Mauritanian epidemic. *Res Virol 140*: 155-64.

Schmaljohn CS 1996. Bunyaviridae: the viruses and their replication, p. 1447-1471. In Fields BN, Knipe DM, Howley PM (eds), *Fields Virology*, 3rd ed., Raven Press, New York.

Smithburn KC 1949. Rift Valley fever: the neurotropic adaptation of the virus and the experimental use of this modified virus as a vaccine. *Brit J Exp Path 30*: 1-16.

Vialat P, Muller R, Vu TH, Prehaud C, Bouloy M 1997. Mapping of the mutations present in the genome of Rift Valley fever virus attenuated MP12 strain and their putative role in attenuation. *Virus Research 52*: 43-50.

Wilson ML 1994. Ecology and the epidemiology of disease emergence. *Ann New York Acad Sci 740*: 169-180.

Zeller HG, Fontenille D, Traoré-Laminzana M, Thiongane Y, Digoutte JP 1997. Enzootic activity of Rift Valley fever virus in Senegal. *Am J Trop Med Hyg 56*: 265-272.

Molecular Epidemiology of Human Polyomavirus JC in the Biaka Pygmies and Bantu of Central Africa

Sylvester C Chima+, Caroline F Ryschkewitsch, Gerald L Stoner

Neurotoxicology Section, National Institutes of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA

Polyomavirus JC (JCV) is ubiquitous in humans and causes a chronic demyelinating disease of the central nervous system , progressive multifocal leukoencephalopathy which is common in AIDS. JCV is excreted in urine of 30-70% of adults worldwide. Based on sequence analysis of JCV complete genomes or fragments thereof, JCV can be classified into geographically derived genotypes. Types 1 and 2 are of European and Asian origin respectively while Types 3 and 6 are African in origin. Type 4, a possible recombinant of European and African genotypes (1 and 3) is common in the USA. To delineate the JCV genotypes in an aboriginal African population, random urine samples were collected from the Biaka Pygmies and Bantu from the Central African Republic. There were 43 males and 25 females aged 4-55 years, with an average age of 26 years. After PCR amplification of JCV in urine, products were directly cycle sequenced. Five of 23 Pygmy adults (22%) and four of 20 Bantu adults (20%) were positive for JC viruria. DNA sequence analysis revealed JCV Type 3 (two), Type 6 (two) and one Type 1 variant in Biaka Pygmies. All the Bantu strains were Type 6. Type 3 and 6 strains of JCV are the predominant strains in central Africa. The presence of multiple subtypes of JCV in Biaka Pygmies may be a result of extensive interactions of Pygmies with their African tribal neighbors during their itinerant movements in the equatorial forest.

Key words: polyomavirus - JC virus - genotypes - Pygmies - Bantu - Africa

The dsDNA polyomavirus JC (JCV) is ubiquitous in humans and bears close sequence homology with other species of this genus, BK virus and the simian virus 40. Sero-epidemiologic studies have shown that up to 90% of adults are positive for antibodies to JCV (Walker & Frisque 1986). Infection with JCV is acquired in early childhood possibly via the respiratory tract. This is followed by persistent infection of the kidneys from which JCV is excreted in urine. Studies with polymerase chain reaction (PCR) show that 30-70% of adults worldwide are positive for JC viruria (Agostini et al. 1996, Sugimoto et al. 1997, Shah et al. 1998). JCV has been established as the causative agent in progressive multifocal leukoencephalopathy (PML), a fatal demyelinating disease of the central nervous system (Zurhein & Chou 1965). PML, previously a rare disorder found in immunocompromised patients with hematologic malignancies, is now prevalent in 5-7% of AIDS cases in the USA and Europe (Berger & Concha 1995, Martinez et al. 1995), but in only 0.8% of Brazilian AIDS patients (Chimelli et al. 1992) and 1.5% in West African AIDS cases (Lucas et al. 1993).

The complete genome of prototype JCV (Mad1) from the brain of a patient with PML was sequenced in 1984 (Frisque et al. 1984). The genome consists of a single molecule of dsDNA, 5.1kb in length, which is transcribed bidirectionally from the origin of DNA replication (ori). It codes for the early region proteins, large T and small t antigens which regulate transcription of the late region proteins VP1-3 and agnoprotein. JCV regulatory region can be classified into two major configurations: an "archetype" which is amplified from urine of normal individuals with JC viruria (Yogo et al. 1990) and a "PML type" when sequenced from the brain of patients with PML. PML-type regulatory regions are derived from the archetypal form by unique rearrangements, consisting of deletions and duplications within the JCV promoter/enhancer (Ault & Stoner 1993, Agostini et al. 1997c).

Based on sequence analysis of JCV complete genomes, as well as segments of the VP1 and T antigen genes, JCV can be classified into several geographically based genotypes and subtypes (Ault & Stoner 1992, Agostini et al. 1995, 1997d, Sugimoto et al. 1997). The major genotypes so far described are Type 1, which is of European origin, Type 2, which is Asian, and Types 3 and 6 which are African in origin (Agostini et al. 1995, 1998).

+Corresponding author. Fax: +301-402-1030.
E-mail: chimasc@helix.nih.gov
Received 15 June 1998
Accepted 30 July 1998

Type 4 which appears to be a recombinant of African and European Types (1 and 3)(Agostini et al. 1996), is prevalent within the United States with the highest frequency in African-Americans. A new clade of JCV strains, consisting of three possible subtypes has been identified in Southeast Asia (Ou et al. 1997) (Chima et al. unpublished data).

Biaka Pygmies (singular 'Aka'), are a group of aboriginal peoples in central Africa who live predominantly as hunter-gatherers in the tropical forest and have a shorter stature when compared to other Africans. Genetic studies have identified Pygmies to have distinctive genetic markers which may be described as "ultra-African" (Cavalli-Sforza 1986). The Biaka show a level of admixture with other Africans, with a residual incidence of 18-35% of ancient Pygmy genes (Cavalli-Sforza 1986, Cavalli-Sforza et al. 1994). It is estimated that the differences between Pygmies and their closest African neighbors are great enough to have required at least 10-20,000 years of isolation, considering that gene flow between this two groups occurs at the rate of only 0.7% per generation (Cavalli-Sforza 1986).

The Biaka Pygmies presented in this study are members of the Babenzele clan, the easternmost subgroup of Aka or "Western" Pygmies, who live in the Dzangha-Sangha dense forest reserve on the banks of the Sangha river, below 4°N of the equator in Central African Republic (C.A.R) (Cavalli-Sforza 1986, Sarno 1995).

The Bantu are African agriculturalists who speak a group of related languages and occupy the southern third of Africa starting from their putative origin in the Nigeria-Cameroon border in the west, to the Kenya-coastline in the east and as far south as Port Elizabeth in South Africa (Hrbek et al. 1992). Pygmies and their Bantu neighbors have a symbiotic relationship of mutual interdependence (Turnbull 1986, Bahuchet 1993, Sarno 1995). It is estimated that the Bantu first made contact with Pygmies during the Bantu expansion about 2-3,000 years ago (Cavalli-Sforza 1986, Hrbek et al. 1992). The Bantu villagers presented in this study live in close proximity and interact extensively with the Pygmies. Indeed, the Biaka and other Pygmy tribes speak a form of Bantu or Nilotic language borrowed from their neighbors having lost their own language over a long period of contact with other African tribes. However, ethnologists and linguists can still recognize common language elements between the Biaka in the west and the most genetically ancient and distant Pygmies (Mbuti), who live in the Ituri forest some 800 miles to the east (Bahuchet 1993, Sarno 1995).

It is assumed that JCV, like any good parasite, has co-evolved with its human host. Due to the stable and distinct JCV genotypes which characterize different populations, urinary JCV has been shown to be a valuable tool in tracing human migrations (Agostini et al. 1997d, Sugimoto et al. 1997). To delineate the JCV genotypes circulating among the aboriginal peoples of central Africa, we undertook a study of the genotype profile of JCV excreted in the urine of the Biaka Pygmies and their Bantu neighbors with a view to determine whether unique strains of JCV may be circulating within these remote people and to compare the rates and pattern of JC viruria with other population groups around the world.

MATERIALS AND METHODS

Patients and samples - Single urine samples (5-50 ml), were collected from 33 Biaka Pygmies from the Pygmy settlement of Yandoumbe and 28 Bantu villagers from Amopolo within the Dsangha-Sangha dense forest reserve in Bayanga prefecture C.A.R. Seven additional urine samples were also collected from two female and five male Bantus living in the city of Bangui, C.A.R. There were 43 males and 25 females with an average age of 26 years and a range of 4-55 years. Adults 20 years and older made up 65% of the sample population. Age determination of the Pygmy population utilized educated estimates by an experienced Pygmy nurse practitioner. All subjects included in the study population were healthy volunteers.

DNA extraction - Urine samples (5-15 ml) were centrifuged at 4,300 x g for 10 min and cell pellets were resuspended in phosphate buffered saline (PBS), recentrifuged and the supernatant was discarded. Cells were suspended in 100-200 µl digestion buffer containing 0.2 mg/ml of proteinase K, 50 mM KCl, 10 mM Tris/HCl (pH 8.3), 2.5 mM $MgCl_2$, 10% (wt/vol) gelatin, 0.45% (vol/vol) NP40 and Tween20. After overnight incubation at 55°C in a waterbath, enzyme reactions were stopped by boiling for 10 min. DNA extracts were stored at -70°C until used and 2-10 µl of the extract was used for subsequent PCR.

PCR - Initial tests for JCV were designed to amplify DNA fragments from the VP1 and large T antigen genes. JCV specific primers for the VP1 coding region were JLP-15 &16 which amplify a 215-bp fragment from this region. This DNA fragment provides up to 15 typing sites for differentiating JCV genotypes and subtypes (JLP-15, nucleotides 1710-1734, 5'ACAGTGTGGCCAGAATT CACTACC-3' and JLP-16, nucleotides 1924-1902, 5'-TAAAGCCT CCCCCCCAACAGAAA-3'). A segment of the large T antigen was amplified using the primer pair JTP-5&6 which amplify a 276-bp fragment from the T-antigen encoding the zinc-finger motif. This region is the site of a mutation

changing a glutamine codon to leucine at amino acid 301. This point mutation is characteristic of all African and some Asian strains of JCV so far studied (Agostini et al. 1995, 1997a) (JTP-5 nucleotides, 3621-3642, 5'-CTTTGTTTGGCTGCTA CAGTAT-3' and JTP-6 nucleotides, 3896-3877, 5'-GCCTTAAGGAGC ATGACTTT-3'). The non coding regulatory regions and T-antigen intron were amplified using the primer pairs JRR-25 & 28 and JSP-1 & 2 respectively. JRR -25 & 28 amplify the entire regulatory region (341-bp) including three typing sites to the left of ori for distinguishing Types 1 and 2 strains (JRR-25, nucleotides, 4981-5004 5'-CATGGATTCCTCCCTA TTCAGCA-3' and JRR-28, nucleotides, 291-268 5'-TCACAGAAGCC TTACGTGACAGC-3'). Specific mutations at positions 133 and 217 of the archetypal regulatory region can be used to further characterize African genotypes. Deletion of certain pentanucleotide repeats within the regulatory region has been used to subtype JCV strains in Taiwan (Ou et al. 1997). The JCV specific primers JSP 1&2 amplify a 402-bp fragment from the T-antigen intron which provides additional typing sites for confirming genotype assignments (JSP-1 nucleotides, 4390-4412, 5'-ACCAGGATTCCCA CTCATCTGT-3' and JSP-2 nucleotides, 4791-4769, 5'-GTTGCTCA TCAGCCTGATTTTG-3').

Following an initial heating at 94°C for 1.5 min (hot start), the 50-cycle, two-step PCR program include 1 min for annealing and elongation at 63°C, denaturation at 94°C for 1 min and extension at 72°C for 1 min. After a final extension for 10 min reactions were terminated at 4°C. PCRs were performed using UlTma DNA polymerase with 3'-5' proofreading activity (Perkin Elmer Cetus) in a standard buffer containing 1.5 mM $MgCl_2$.

Cycle sequencing - Gel-purified PCR products were sequenced directly using the Excel Kit (Epicentre Technologies, Madison, WI) with the same primers used for DNA amplification end-labeled with ^{33}P-ATP (Amersham, Arlington Heights, IL). Initial denaturation at 95°C was followed by 30 cycles of 30 sec at 95°C for denaturation and 1 min at 63°C for annealing and elongation. Products were electrophoresed on a 6% polyacrylamide gel containing 50% urea. Gels were fixed with 12% methanol and 10% acetic acid, transferred to 3MM chromatography paper, dried under vacuum, then exposed to X-ray film for 12-48 hr.

JCV genotypes were identified as previously described (Ault & Stoner 1992, Agostini et al. 1995, 1997b, 1997e, 1998). Sequence relationships were analyzed with GCG programs, Unix version 8 (Genetics Computer Group, Madison, WI). Primer design was assisted by the OLIGO program version 5.0 (NBI, Plymouth, MN).

Reference sequences - The following are GenBank accession numbers for JCV sequences referred to in this work: JCV archetypal regulatory region JCV(CY) M35834 (Yogo et al. 1990); JCV coding region JCV(Mad-1), J02227 (Frisque et al. 1984); JCV Type 6 coding and regulatory regions, AF015537 and AF015538 (Agostini et al. 1998); JCV Type 3 strains #309, U73178, #311, U73501 (Agostini et al. 1997a); JCV strain#123, subtype 1B, AF015527 (Agostini et al. 1997b).

RESULTS

The age and gender of the Biaka and Bantu adults tested for JC viruria is given in the Table. Of the 43 adults tested by PCR amplification of the VP1 coding region, 22% (5 of 23) Pygmies and 20% (4 of 20) Bantus were shown to excrete the virus in urine. Overall, males had a higher excretion rate than females, seven out of 27 (26%) compared with two out of 16 (13%). None of the 24 children and adolescents aged 18 years or younger included in the sample population were positive for JC viruria. One of seven samples collected from Bantus in the city of Bangui was positive. This strain, L1081, was obtained from the urine of a 47-year old Cameroonian of the Bemoun tribe long domiciled in C.A.R.

JCV coding regions - The JCV genotypes excreted by the nine adults were further analyzed by direct cycle sequencing of the JLP-15 & 16 amplified fragments from both directions. Within this

TABLE

Age and gender of Pygmy and Bantu adults screened for JC viruria

Cohort	Gender	No. adults	Age range (years)	No. positives	% positives
Pygmy	M	15	25-55	3	20
	F	8	30-55	2	25
	Total	23		5	22
Bantu	M	12	21-55	4	33
	F	8	22-40	0	0
	Total	20		4	20

fragment up to 18 typing sites have been identified for differentiating JCV genotypes and subtypes. Fourteen of these sites are illustrated in Fig. 1. JCV Type 6 can be clearly distinguished from both Types 1 and 3 at positions 1790 and 1837. Type 1 strains can be separated from both Types 3 and 6 at position 1771, while the two subtypes of Type 1, (1A and 1B) can be differentiated from each other at positions 1843 and 1850.

Analysis of the JCV strains from Pygmies yielded three different types of JCV from five positive samples. These were two Type 3 strains, one Type 1 and two Type 6 strains. One of the Type 3 strains (L1059) showed identical sequence in the VP1 fragment to the DNA sequence of strain #309 previously amplified from the urine of an African from Mara region in Tanzania (Agostini et al. 1995). The other Type 3 strain (L1066) showed partial sequence homology with #311(Type 3B), previously sequenced from an African-American, but differed from this strain at position 1870 where deoxyadenosine was inserted in place of deoxyguanosine. The latter strain was therefore termed a variant of Type 3B pending analysis of the complete genome. Strain L1132, from a Biaka Pygmy showed very close sequence homology in the VP1 fragment when compared to a Type 1B strain, #123, sequenced from a Caucasian (Agostini et al. 1997b). However this Aka strain had a distinct point mutation at position 1830, where deoxythymidine (T) was replaced by a 'G'. This mutation caused a change in the codon for amino acid inserted at this position from valine to glycine. This point mutation at position 1830 of Aka strain L1132 has not been described previously in any Type 1 strains (Agostini et al. 1997b). Both Type 6 strains sequenced from Aka were identical with the previously reported Type 6 sequence (#601). A total of four JCV strains were sequenced from the Bantu. These four strains when analyzed showed exact sequence homology in the JLP-15 and 16 amplified fragments when compared to strain #601, sequenced from the brain of an African-American patient with PML. The Bantu Type 6 strains were also identical to the Aka Type 6 (Fig. 1).

Strain No.	Ethnicity	1753	1771	1786	1790	1795	1804	1805	1818	1830	1837	1843	1850	1869	1870	JCV
Mad1	Caucasian	A	C	G	T	A	T	A	G	T	T	G	A	G	G	Type 1A
#123	Caucasian	A	C	G	T	A	T	A	G	T	T	T	G	G	G	Type 1B
L1132	Aka	A	C	G	T	A	T	A	G	G	T	T	G	G	G	Type 1B
#309	Tanzanian	T	A	A	T	A	C	G	C	T	T	T	A	C	A	Type 3A
L1059	Aka	T	A	A	T	A	C	G	C	T	T	T	A	C	A	Type 3A
#311	AfAm	T	A	G	T	A	T	G	C	T	T	T	A	C	G	Type 3B
L1066	Aka	T	A	G	T	A	T	G	C	T	T	T	A	C	A	Type 3B
#601	AfAm	A	A	A	C	A	T	A	C	T	C	T	A	G	G	Type 6
L1069	Aka	A	A	A	C	A	T	A	C	T	C	T	A	G	G	Type 6
L1076	Aka	A	A	A	C	A	T	A	C	T	C	T	A	G	G	Type 6
L1044	Bantu	A	A	A	C	A	T	A	C	T	C	T	A	G	G	Type 6
L1052	Bantu	A	A	A	C	A	T	A	C	T	C	T	A	G	G	Type 6
L1081	Bantu	A	A	A	C	A	T	A	C	T	C	T	A	G	G	Type 6
L1138	Bantu	A	A	A	C	A	T	A	C	T	C	T	A	G	G	Type 6

Fig. 1: typing sites within the JLP-15& 16 amplified fragments of the VP1 gene. Bantu and Pygmy strains are compared to JCV Mad1 sequence and strains #123 (Type 1B) (Agostini et al. 1997b), #309 (Type 3A) from Tanzania, #311 (Type 3B) and # 601 (Type 6) from African-Americans (Agostini et al. 1997a, 1998). L1132 shows a point mutation at nucleotide 1830. L1066 shows similarity with Type 3B nucleotides at positions 1786 and 1804 (solid frame), while it resembles Type 3A at position 1870 (broken frame). Numbering is based on the sequence of JCV Mad1 (Frisque et al. 1984).

A 276-bp fragment was sequenced from the large T antigen of six JCV strains (three Aka and three Bantu) using the Primer pair JTP- 5 and 6. This T antigen fragment encodes the zinc finger motif. A specific point mutation in this fragment characterizes all African strains of JCV so far described and some Asian strains. This mutation is a non-conservative nucleotide base substitution at position 3768 from 'T' to 'A', causing a change in the amino acid coded from hydrophilic glutamine to hydrophobic leucine (Agostini et al. 1997a). The six Bantu and Pygmy strains amplified from the T-antigen zinc finger region showed a mutation at position 3768 (Fig. 2). Typing sites within this fragment confirm strain L1059 as a Type 3 strain and strains L1052, L1069, L1076, L1081 and L1138 as Type 6 strains.

JCV noncoding regions - Noncoding regulatory regions of six JCV strains from Bantus and Pygmies were sequenced by the primers JRR-25 and 28 from both directions. The DNA sequence was compared to the consensus archetypal sequence of Type 1 (Agostini et al. 1996) and a Type 3 regulatory region sequence #309 from an Tanzanian (Agostini et al. 1997a). The Aka Type 3 strain (L1059) showed sequence identity with #309 including a point mutation at position 133 where 'C' is characteristic of all Type 3 strains. Four Type 6 strains from Bantus and Pygmies, (L1052, L1069, L1076, and L1138) all showed an archetypal configuration without deletions. Strains L1081 (Type 6, Bantu) and L1059 (Type 3, Aka) both show a 10-bp deletion at nucleotides (51-60), just preceding the first NF1 site (Fig. 3). The deletion at this site is identical to those observed in strains #307 and #309 from Tanzania (Agostini et al. 1995, 1997a). All the Type 6 strains and the single Type 3 strain were characterized by the nucleotide "G" at position 217, however only the Type 3 strain showed deoxycytosine at position 133 of the regulatory region.

A 402-bp fragment was amplified from the noncoding T-antigen intron using the primers JSP-1 and 2. This fragment provides up to 15 additional typing sites for confirmation of JCV types and subtypes from the coding region sequences. Seven JCV strains were amplified from this fragment in the Pygmy and Bantu cohorts. Cycle sequencing confirmed the previous type assignments from the VP1 gene. L1044 (Bantu, Type 6) showed two nucleotide mutations at positions 4562 and 4648 while L1059 (Aka, Type 3) showed a single mutation at position 4435 (Fig. 4). The significance of these point mutations is unknown since the primary function of introns is to be spliced out prior to protein translation.

DISCUSSION

This study delineates the genotype profile of JCV strains circulating among the Biaka Pygmies and Bantu from Bayanga prefecture of C.A.R. This aboriginal African population excretes JCV in urine at a lower rate (21%) when compared to rates of excretion in urban populations in the United States

JTP 5&6

JCV	3680	3710	3722	3743	3768	3770	3809	3830	3836	3848
Mad 1	A	T	G	C	T	T	A	T	G	A
Type 3	A	C	G	C	A	T	G	G	A	T
Type 6	A	C	A	C	A	T	A	T	G	A
L1052	A	C	A	C	A	T	A	T	G	A
L1059	A	C	G	C	A	T	G	G	A	T
L1069	A	C	A	C	A	T	A	T	G	A
L1076	A	C	A	C	A	T	A	T	G	A
L1081	A	C	A	C	A	T	A	T	G	A
L1138	A	C	A	C	A	T	A	T	G	A

Fig. 2: typing sites within the JTP-5&6 amplified fragment of large T antigen including the zinc finger motif. Position 3768 (frame) shows site of nucleotide mutation from "T" to "A" in all African genotypes including Bantu and Pygmy strains when compared to JCV Mad1.

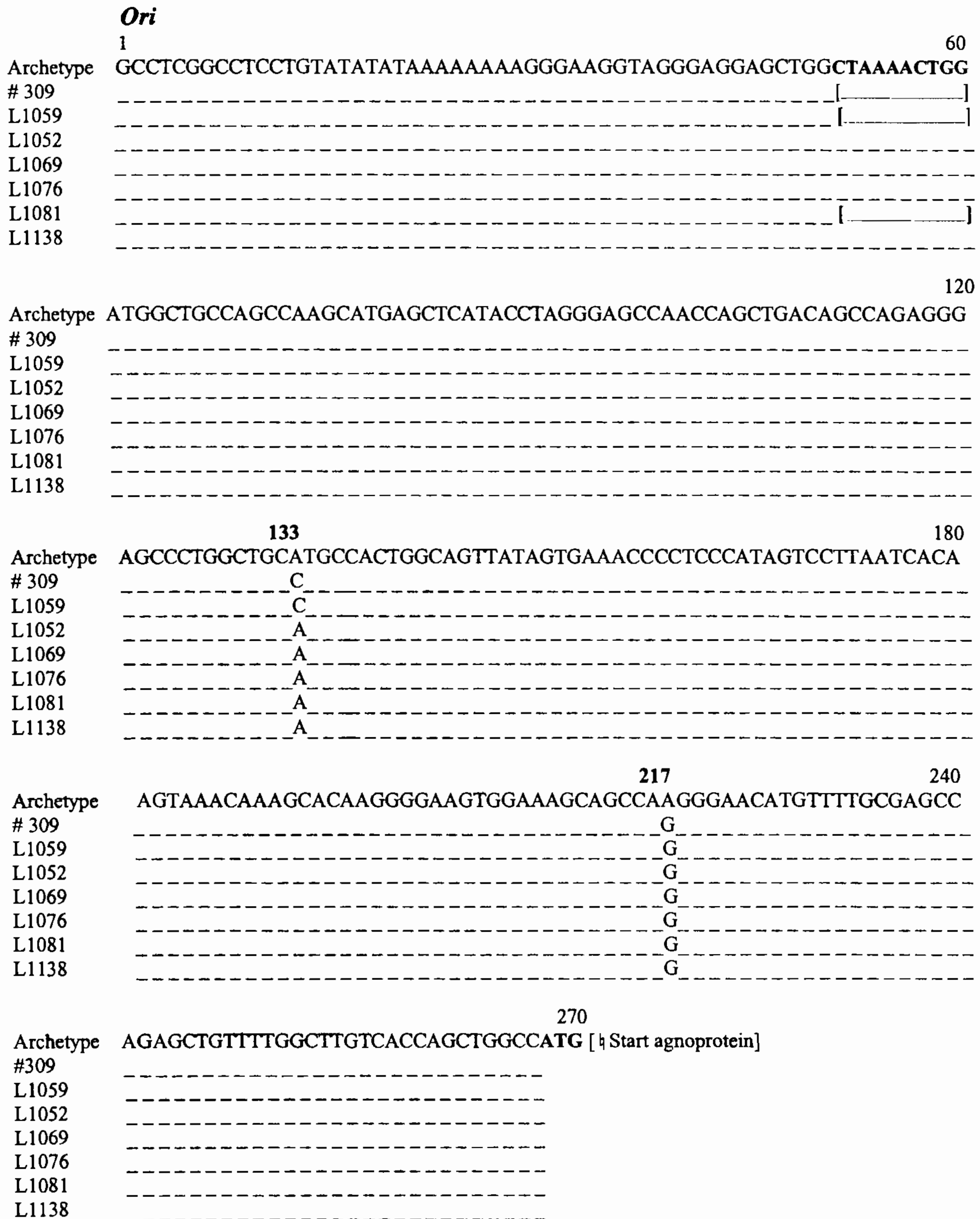


Fig. 3: regulatory region sequences amplified from Pygmy and Bantu strains is compared to the consensus archetypal regulatory region of Type 1 (Agostini et al. 1996) and #309 from Tanzania. Dashed lines denote uniformity with the consensus archetypal sequence. Solid lines show areas of nucleotide deletion initially observed in strains #307 and #309 (Agostini et al. 1995, 1997a) and now found in L1059 from a Biaka Pygmy and L1081 from a Bantu. At position 133, "A" is replaced by "C" in all Type 3 strains. At position 217, both Type 3 and Type 6 strains substitute deoxyguanosine for deoxyadenosine. Numbering is based on archetypal numbering of strain CY (Yogo et al. 1990).

JSP 1&2

JCV	4430	4434	4435	4438	4442	4466	4477	4481	4484	4519	4522	4534	4562	4648	4652	4714	4767
Mad1	A	A	A	C	T	G	G	A	A	G	G	T	G	G	C	G	C
Type 3	C	A	A	A	T	G	C	G	T	G	A	C	G	G	C	A	T
#601	A	A	A	A	T	A	C	A	T	A	G	C	G	G	T	G	C
L1044	A	A	A	A	T	A	G	A	T	G	G	C	**T**	**A**	T	G	C
L1052	A	A	A	A	A	A	G	A	T	G	G	C	G	G	T	G	C
L1059	C	A	**G**	A	T	G	C	G	T	G	A	C	G	G	C	A	T
L1069	A	A	A	A	A	A	G	A	T	G	G	C	G	G	T	G	C
L1076	A	A	A	A	A	A	G	A	T	G	G	C	G	G	T	G	C
L1081	A	A	A	A	A	A	G	A	T	G	G	C	G	G	T	G	C
L1138	A	A	A	A	A	A	G	A	T	G	G	C	G	G	T	G	C

Fig. 4: the JSP-1&2 amplified fragment of the T antigen intron further confirm genotype assignments from the VP1 and large T antigen genes. Typing in this region is compared to the consensus sequence of Type 3 (Agostini et al. 1997a), strain #601 (Agostini et al. 1998) and Mad1. Framed sets denote sites of specific point mutations in L1044 and L1059 from Biaka Pygmies. Numbering is based on Mad1 sequence.

(41%) (Agostini et al. 1996) and Europe (Stoner et al. 1998a). Native American tribes in the United States and the Pacific Islands show a rate of JC virus excretion in urine (65%) (Agostini et al. 1997d), which is three times the rate observed in this African cohort. However the rate of excretion among the Bantu and Pygmies are somewhat closer to a reported incidence rate of 30% in HIV positive patients from the Mara region of northwest Tanzania (Agostini et al. 1995). The reasons for the differences in rates of JCV virus excretion in different populations is not yet explained. However, it may be related in part to the difference in age of various sample populations. Studies in Caucasians and African-American cohorts within the United States have shown that the rate of JC virus excretion in urine rises dramatically in the fifth decade of life (Agostini et al. 1996), (Chima, unpublished observations). It therefore follows that sample populations with older age groups are more likely to yield a higher rate of JC viruria. The African cohort studied here had only three adults estimated to be aged 50 years or older.

Analysis of the JCV strains from Pygmy urine revealed four different subtypes from the five positive cases. These were two Type 3 strains (one 3A and one 3B variant), two Type 6 and one Type 1B variant. The Type 3A strain showed close identity with Type 3 strains previously reported among Nilotic Africans of the Luo tribe from the Mara region of Tanzania. The Type 3B strain showed a similar sequence to that recently found in an African-American (strain A179) (Chima, unpublished data). This is a variant of strain #311 also found in an African-American with an 'A' to 'G' substitution at position 1870 of the VP1 gene. The two Type 6 strains were identical to those sequenced from the urine of the Bantu in this study.

JCV Type 6 was first sequenced from the brain of an African-American patient with PML (Agostini et al. 1998). This was later identified as a new subtype of JCV when similar strains were sequenced from the urine of Africans from Ghana (Guo et al. 1996). Type 6 strains have also been sequenced from the brains of AIDS patients with PML from the Ivory Coast (Stoner et al. 1998b) as well as the urine of an immunocompetent individual from Sierra Leone (Chima, unpublished data). The four JCV strains excreted in the urine of Bantus reported here are Type 6. Of the four Bantu strains, (L1081) showed a 10-bp deletion in the regulatory region sequence similar to that found

in #309 from Tanzania and L1059 in Pygmies. However, L1059 also displays another marker of Type 3 strains, i.e., deoxycytosine at position 133 of the archetypal regulatory region. It is more likely therefore, that these two strains arose independently of each other rather than as a result of viral recombination. We can hypothesize that the two African genotypes of JCV (Types 3 and 6) may have coevolved, independently of each other, in their respective African hosts. All genotype studies on JCV in Africans so far have shown that both Type 3 and 6 strains can be found in West and Central Africa (Guo et al. 1996, Sugimoto et al. 1997, Stoner et al. 1998b), while Type 3 is the only genotype so far described from East Africa (Agostini et al. 1995).

Archeological and linguistic data have shown that the Biaka Pygmies migrated to their present location from a region north of the Ituri around the southern Sudan, first to northern Zaire and then in a northwest direction to their present location in the southwest tip of C.A.R. around the Sangha river (Cavalli-Sforza 1986, Bahuchet 1993). The putative site of Biaka Pygmy origin around the southern Sudan is closer to the region occupied by previously studied Africans from northwest region of Tanzania. The latter population are in part Nilotics of the Luo tribe (Agostini et al. 1995). This group excrete Type 3 JCV strains similar to those found in Biaka Pygmies. The Bantus on the other hand are migratory farmers thought to have come into contact with the Pygmies about 2000 years ago during the Bantu expansion from West Africa (Cavalli-Sforza 1986, Hrbek et al. 1992). Archeologists and historians estimate that during the second stream of the Bantu expansion, there was a migration along the banks of the Sangha river into central Africa (Hrbek et al. 1992). It is therefore likely that Bantu descendants of the first immigrants still occupy the present location and carry JCV strains transmitted from their parents. Due to the close interaction between the Pygmies and their Bantu or Nilotic neighbors in equatorial Africa, it may be speculated that Type 6 strains were transmitted to the Biaka during their later interactions with Bantus while the Type 3 strains were brought along during their migration from southern Sudan and East Africa.

A Type 1B variant of JCV was sequenced from the urine of a 55 year old female Pygmy. Type 1 strains are generally characteristic of Europeans. This Aka strain bears a unique mutation at position 1830 not previously reported in Type 1 strains of JCV (Agostini et al. 1997b, Stoner et al. 1998a). The significance of this Type 1 strain is unknown although in another study, it has been reported that a pocket of the European subtype of JCV was found in Bangui, C.A.R. (Sugimoto et al. 1997). Analysis of the complete genome of the Aka Type 1B variant and identification of more JCV strains with similar mutations will facilitate characterization of this subtype. It is possible that on analysis of the complete genome, this strain may represent a unique subtype of JCV different from Type 1 strains

We conclude that human polyomavirus JCV is excreted in the urine of Biaka Pygmies and Bantus of central Africa, though at a lower rate than that observed in other population groups. This study confirms Types 3 and 6 as the predominant genotypes of JCV in central Africa. The finding of four different subtypes of JCV in the urine of Biaka Pygmies may be explained by the extensive interactions of Pygmies with their various African tribal neighbors over a long period of time, as they moved from place to place in the equatorial forest.

ACKNOWLEDGMENTS

To Hansjurgen T Agostini for initial studies on African genotypes of JC virus. To the entire staff of the World Wildlife Fund in Bangui and Bayanga for their kind hospitality and assistance throughout our stay in the Central African Republic.

REFERENCES

Agostini HT, Brubaker GR, Shao J, Levin A, Ryskcwitsch CF, Blattner WA, Stoner GL 1995. BK virus and a new type of JC virus excreted by HIV-1 positive patients in rural Tanzania. *Arch Virol 140*: 1919-1934.

Agostini HT, Ryschkewitsch CF, Stoner GL 1996. Genotype profile of humam polyomavirus JC excreted in urine of immunocompetent individuals. *J Clin Microbiol 34*: 159-164.

Agostini HT, Ryschkewitsch CF, Stoner GL 1998 . The complete genome of JC Virus Type 6 from the brain of a African-American with progressive multifocal leukoencephalopathy (PML). *J Hum Virol*: in press.

Agostini HT, Ryschkewitsch CF, Brubaker GR, Shao J, Stoner GL 1997a. Five complete genomes of JC virus Type 3 from Africans and African Americans *Arch Virol 142*: 637-655.

Agostini HT, Ryschkewitsch CF, Singer CF, Stoner GL 1997b. JC virus Type 1 has multiple subtypes: three new complete genomes. *J Gen Virol 79*: 801-805.

Agostini HT, Ryschkewitsch CF, Singer EJ, Stoner GL 1997c. JC virus regulatory region rearrangements and genotypes in progressive multifocal leukoencephalopathy: two independent aspects of virus variation. *J Gen Virol 78*: 659-664.

Agostini HT, Ryschkewitsch CF, Yanagihara R, Davis V, Stoner GL 1997d. Asian genotypes of JC virus (JCV) in Native Americans and in a Pacific Island population: markers of human evolution and migration, *Proc Natl Acad Sci US* 94: 14542-14546.

Agostini HT, Shishido Y, Ryschewitsch CF, Stoner GL 1997e. JC Virus Type 2: definition of subtypes based on analysis of ten complete genomes. *J Gen Virol*: in press.

Ault GS, Stoner GL, 1992 . Two major types of JC virus defined in progressive multifocal leukoencephalopathy brain by early and late coding region DNA sequence *J Gen Virol 73*: 2669-2678.

Ault GS, Stoner GL 1993. Human polyomavirus JC promoter/enhancer rearrangement patterns from progressive multifocal leukoencephalopathy brain are unique derivatives of a single archetypal structure. *J Gen Virol 74*: 1499-1507.

Bahuchet S 1993. History of the inhabitants of the Central African Rain Forest: Perspectives from comparative linguistics, p. 37-54. In CM Hladik, A Hladik, O Linares, H Pagezy, A Semple & M Hadley (eds), *Tropical Forests, People and Food*, UNESCO, Paris.

Berger JR, Concha M 1995. Progressive multifocal leukoencephalopathy: the evolution of a disease once considered rare. *J Neurovirol 1*: 5-18.

Cavalli-Sforza LL 1986. *African Pygmies*, Academic Press, Orlando.

Cavalli-Sforza LL, Menozzi P, Piazza A 1994. Africa, p. 159-194. In Cavalli-Sforza LL, Menozzi P, Piazza A (eds), *The History and Geography of Human Genes*, Princeton University Press, Princeton.

Chimelli L, Rosemberg S, Hahn MD, Lopes MBS, Barretto-Netto M 1992. Pathology of the central nervous system in patients infected with the human immunodeficiency virus (HIV): a report of 252 autopsy cases from Brazil. *Neuropathol Appl Neurobiol 18*: 478-488.

Frisque RJ, Bream GL, Cannella MT 1984. Human polyomavirus JC virus genome. *J Virol 51*: 458-469.

Guo J, Kitamura T, Ebihara H, Sugimoto C, Kunitake T, Takehisa J, Na YQ, Al-Ahdal MN, Hallin A, Kawabe K, Taguchi F, Yogo Y 1996. Geographical distribution of the human polyomavirus JC virus type A and B and isolation of a new type from Ghana. *J Gen Virol 77*: 919-927.

Lwango-Lunyiigo S, Vansina J 1992. The Bantu-speaking peoples and their expansion, p. 75-85. In I Hrbek, *General History of Africa, Vol. 111, Africa from the Seventh to the Eleventh Century*, UNESCO, Paris.

Lucas SB, Hounnou A, Peacock C, Beaumel A, Djomand G, N'gbichi J-M, Yeboue K, Honde M, Diomande M, Giordano C, Doorly R, Brattegaard K, Kestens L, Smithwick R, Kadio A, Ezani N, Yapi A, De Cock KM 1993. The mortality and pathology of HIV infection in a West African city. *AIDS 7*: 1569-1579.

Martinez AJ, Sell M, Mitrovics T, Stoltenburg-Didinger G, Inglesias-Rojas JR, Giraldo-Velasquez MA, Gosztonyi G, Schneider V, Cervos-Navarro J 1995. The neuropathology and epidemiology of AIDS. A Berlin experience. A review of 200 cases. *Path Res Pract 191*: 427-449.

Ou W, Tsai R, Wang M, Fung C, Hseu T, Chang D, 1997. Genomic Cloning and sequence analysis of Taiwan-3 Human Polyomavirus JC Virus. *J Formos Med Assoc 96*: 511-516.

Sarno L 1995. *Bayaka: the Extraordinary Music of the Babenzele Pygmies*, Ellipsis Arts, New York, 66 pp.

Shah KV, Daniel RW, Strickler HD, Goedert JJ 1998. Investigation of human urine for genomic sequences of the primate polyomaviruses simian virus 40, BK virus, and JC virus. *J Infect Dis 176*: 1618-1621.

Stoner GL, Agostini HT, Ryschkewitsch CF, Komoly S 1998a. JC virus excreted by multiple sclerosis patients and paired controls from Hungary. *Multiple Sclerosis 4*: 45-48.

Stoner GL, Agostini HT, Ryschewitsch CF, Mazlo M, Gullotta F, Wamukota W, Lucas S 1998b. Two cases of progressive multifocal leukoencephalopathy (PML) due to JC virus: detection of JCV Type 3 in a Gambian AIDS patient. *J Med Microbiol 47*: 1-10.

Sugimoto C, Kitamura T, Guo J, Al-Ahdal MN, Schelnukov SN, Otova B, Ondrejka P, Chollet JY, El-Safi S, Ettayebi M, Gresenguet G, Kocagoz T, Chaiyarasamee S, Thant KZ, Thein S, Moe K, Kobayashi N, Taguchi F, Yogo Y 1997. Typing urinary JC virus DNA offers a novel means of tracing human migrations. *Proc Natl Acad Sci USA 94*: 9191-9196.

Turnbull CM 1986. Survival factors among Mbuti and other hunters of the Equatorial Rain Forest, p. 103-123. In Cavalli-Sforza LL, *African Pygmies*, Academic Press, Orlando.

Walker DL, Frisque RJ 1986. The biology and molecular biology of JC virus, p. 327-377. In NP Salzman *The Papovaviridae*, Vol. 1, *The Polyomaviruses*, Plenum Press, New York.

Yogo Y, Kitamura T, Sugimoto C, Ueki T, Aso Y, Hara K, Taguchi F 1990. Isolation of a possible archetypal JC virus DNA sequence from non immmunocompromised individuals. *J Virol 64*: 3139-3143.

Zurhein GM, Chou SM 1965. Particles resembling papova viruses in human cerebral demyelinating disease. *Science 148*: 1477-1479.

RESEARCH NOTE

Molecular Epidemiology of Den-2 Virus in Brazil

MP Miagostovich/+, RMR Nogueira, HG Schatzmayr, RS Lanciotti*

Laboratório de Flavivirus, Departamento de Virologia, Instituto Oswaldo Cruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil *Centers for Disease Control and Prevention, CDC, Fort Collins, CO, USA

Key words: dengue virus type 2 - sequencing - Brazil

Dengue (DEN) viruses belong to the family *Flaviviridae*, genus *Flavivirus*, and occur as four antigenically related, but distinct serotypes designated DEN-1, 2, 3 and 4 (EG Westaway et al. 1985 *Intervirology 24*: 183-192). The viruses are characterized by a single strand of RNA associated with a core protein, in a nucleocapside surrounded by a lipid envelope. The genoma consist of a single open reading frame coding for core protein (C), precursor of the membrane protein (prM) and envelope (E) structural proteins, followed by the non structural proteins NS1, NS2a, NS2b, NS3, NS4a, NS4b and NS5 (FZ Heinz & JT Roehring 1990 p. 289-305. In MHV Van Regenmorte, AR Neurath (eds), *Immunochemistry of Viruses, II. The Basis for Serodiagnosis and Vaccines*, Elsevier).

The genetic variation among DEN viruses has been demonstrated by numerous methods including oligonucleotides fingerprinting, restriction enzymes, primer extension sequencing and nucleotide sequences from different fragments of genoma (R Rico-Hesse 1990 *Virology 174*: 479-493, JA Lewis et al. 1993 *Virology 197*: 216-224, V Deubel et al. 1993 *Arch Virol 129*: 197-210, V Vorndam et al. 1994 *Arch Virol 136*: 191-196, DW Trent et al. 1989 *Virology 172*: 523-535, RS Lanciotti et al. 1994 *J Gen Virol 75*: 65-75, E Chungue et al. 1995 *J Gen Virol 76*: 1877-1884, KZ Thant et al. 1995 *Microbiol Immunol 39*: 581-590).

Since intra-serotypic antigenic variations (genetic subtypes) could be associated with severe disease it is important to monitor the distribution and eventual introduction of new genotypes of existing serotypes into areas where dengue activity are troublesome (Vorndam et al. 1994 *loc. cit.*).

In this report we sequenced the E fragment from geographically and temporally distinct DEN-2 viruses isolated in Brazil during 1990-1995, in order to investigate the genetic subtype distribution of this serotype virus in the country.

DEN-2 viruses analyzed in this study were obtained from the collection of the Laboratory of Flavivirus, Department of Virology, IOC, Fiocruz. These strains were isolated from sera by inoculation into *Aedes albopictus* clone C6/36 cell line (A Igarashi 1978 *J Gen Virol 40*: 531-544) and were identified by immunofluorescence using type-specific monoclonal antibodies. Virus seeds were amplified once by inoculation into C6/36 (DJ Gubler et al. 1984 *Am J Trop Med Hyg 33*: 158-165).

Viral RNA was extracted from infected C6/36 cells by using the acid-guanidin isothiocyanate procedure previously described (RS Lanciotti et al. 1992 *J Clin Microbiol 30*: 545-551). Oligonucleotide primers used in the amplification and sequencing protocols were designed with the aid of the oligo program (National Bioscience Inc., Plymouth, MN).

Nucleotides from positions 1685 to 2504 coding for the fragment of E gene were amplified using RT-PCR. The RT reaction was performed in 4 µl of 5X RT reaction buffer (BRL), 4 µl of water, 2 µl of 0.1M DTT, 5 µl of 25 µM dNTP's, 0.2 µl of Rnasin (40U/µl), 2µl of 10 mM downstream primer D2CP2504 (5' GGGGATTCTGGTTGGAACTT ATATTGTTCTGTCC 3'), 2 µl of RNA and 1 µl of 200 U Superscript RT (Gibco). The reaction was incubated at 50°C for 10 min then 50 min at 45°C. A PCR amplification was followed by adding 10 µl of RT reaction to 90 µl of PCR reaction mix (74 µl of water, 9 µl of 10X C buffer, 5 µl of 25 mM dNTP, 2µl of 10 mM upstream primer D2P1685 (5'CTAGGATCTCAAGAAGGAGC AATGCA 3') and 0.5 µl Taq. The DNA molecules were denaturated at 94°C for 4 min and subjected to a 35 amplification cycles (94°C for 1 min, 55°C for 1 min, 72°C for 8 min) and to one of 72°C for 10 min.

After an eletrophoresis on a 1% agarose gel, the amplified DNA bands were excised and purified by using the Bio 101 Gene-Clean kit. Purified DNAs were then sequenced by using the following primers: D2P1685- 5'CTAGGATCTCAA GAAGGAGCAATGCA 3'; P760 - 5' GGATCA CAAGAAGGAGCCATGCA 3'; CP1171 - 5' ATGGAGCTTCCTTTCTTCTTGAACCA 3'; CP1234 - 5' CCAAAGTCCCAGGCTGTGTCTC CCAGAATGGCCAT 3'. The sequencing reaction was performed by using the Taq DyeDeoxy Terminator Cycle Sequencing Kit (Applied Biosystem, Inc., USA). Cycle sequencing parameters were exactly as described in the manufactures protocol.

Financial support: CNPq and Fundação Banco do Brasil.
+Corresponding author. Fax: + 55-21-270.6397. E-mail: marizepm@gene.dbbm.fiocruz.br
Received 15 June 1998
Accepted 30 July 1998

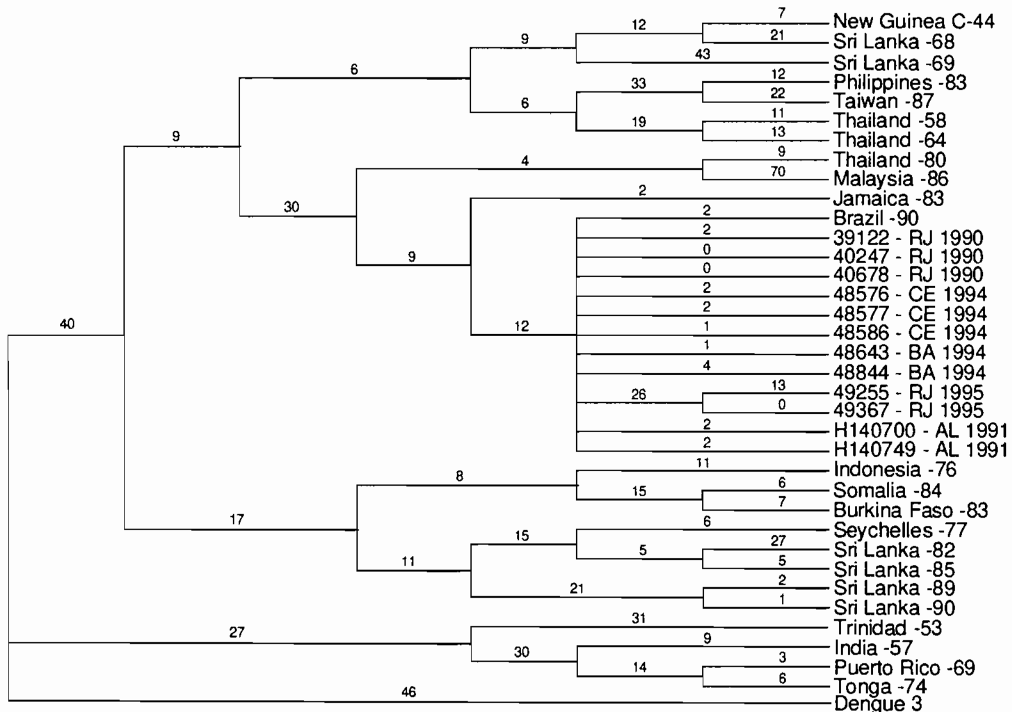
The overlapping nucleic acid sequences obtained from individual sequencing reactions were combined for analysis and edited using the DNASTAR program (Madison, WI). The DEN-2 virus nucleic acid sequences were then aligned with each other, and with DEN-2 envelope sequences obtained from GENEBANK, using the multiple sequence alignment algorithm CLUSTAL (D Higgins, Heidelberg, Germany). Phylogenetic trees were reconstructed from the aligned nucleic acid sequences using algorithms based upon parsimony (program PAUP, D Swofford, Champaign, IL).

The comparison of our results with the phylogram generated by the sequencing of the entire E gene (Lewis et al. 1993 *loc. cit.*) showed that all the isolates belong to subtype III (Figure). The results confirmed the asiatic origin of DEN-2 strains isolated in the State of Rio de Janeiro previously demonstrated by Rico-Hesse (*1990 loc. cit.*) and Lewis et al. (1993 *loc. cit.*). The circulation of the same genotype in all areas studied demonstrated the dispersion of DEN-2 virus from Rio de Janeiro to the other states of the country. The subtype III has been refered to have a greater potential to cause severe disease causing concern in those areas in which high rates of antibody to DEN-1 and DEN-4 viruses predispose populations to severe disease (Vorndam et al. 1994 *loc. cit.*).

In Brazil, the increasing incidence of DHF/DSS was associated with the introduction of the DEN-2 viruses in the states of Rio de Janeiro, Ceará and recently in the State of Rio Grande do Norte (RMR Nogueira et al. 1993 *Epidemiol Infect 111:* 163-170, RV Souza et al. 1995 *Mem Inst Oswaldo Cruz 90*: 345-346, PFC Vasconcelos et al. 1995 *Rev Inst Med Trop São Paulo 37*: 253-255, SMO Zagne et al. 1994 *Trans R Soc Trop Med Hyg 88*: 677-679) after a period of high DEN-1 virus activity. In the states of Bahia and Espírito Santo, where DEN-2 virus was responsible initially for primary infections, signs and symptoms of classic dengue fever were observed (RMR Nogueira et al. 1995 *Rev Inst Med Trop São Paulo 37*: 507-510). In those states a higher percentage of exhantema and pruritus were observed when compared with signs and symptoms due to DEN-1 primary infection during 1986 in Rio de Janeiro.

Recently, R Rico-Hesse et al. (1997 *Virology 230*: 1-8) demonstrated the direct association between the introduction of southeast Asian DEN-2 viruses severe disease in America and showed a circulation of a new subtype responsible for DHF epidemics in Mexico and Venezuela, in 1995. This data point out the need to continue molecular epidemiological studies in dengue endemic areas in order to monitor the introduction of a new subtype and the impact of it over the population.

Acknowledgement: to Dr J Chang for supplying the primers.



Phylogram generated by parsimony analysis of nucleic acid sequences from the prM/M and E genes of 12 DEN-2 Brazilian strains and 24 DEN-2 viruses obtained from Genebank.

Antimalarial Drug Resistance: Surveillance and Molecular Methods for National Malaria Control Programmes

Umberto D'Alessandro

Prince Leopold Institute of Tropical Medicine, Antwerp, Belgium

National malaria control programmes have the responsibility to develop a policy for malaria disease management based on a set of defined criteria as efficacy, side effects, costs and compliance. These will fluctuate over time and national guidelines will require periodic re-assessment and revision. Changing a drug policy is a major undertaking that can take several years before being fully operational. The standard methods on which a decision can be taken are the in vivo *and the* in vitro *tests. The latter allow a quantitative measurement of the drug response and the assessment of several drugs at once. However, in terms of drug policy change its results might be difficult to interpret although they may be used as an early warning system for 2nd or 3rd line drugs. The new WHO 14-days* in vivo *test addresses mainly the problem of treatment failure and of haematological parameters changes in sick children. It gives valuable information on whether a drug still 'works'. None of these methods are well suited for large-scale studies. Molecular methods based on detection of mutations in parasite molecules targeted by antimalarial drugs could be attractive tools for surveillance. However, their relationship with* in vivo *test results needs to be established.*

Key words: antimalarial drug resistance - *in vivo* test - *in vitro* test - polymerase chain reaction - surveillance

Despite considerable efforts done during this century to eradicate or control it, *Plamodium falciparum* malaria is still the most prevalent and the most devastating disease in the tropics (WHO 1993). In the last decades, its control and treatment has been complicated by the emergence of resistance to widely used antimalarial drugs such as chloroquine. Drug resistance of malaria parasites has been defined as the ability of a parasite strain to multiply or to survive in the presence of concentrations of a drug that normally destroy parasites of the same species or prevent their multiplication. Its dynamics and occurrence are the result of several interactions:

Parasites and drugs - Natural populations of *P. falciparum* are heterogeneous mixtures of individuals with different, genetically determined degrees of drug response. The efficacy of medication will depend on the concentration of the drug in relation to the parasite's sensitivity and the time over which concentrations above this threshold are maintained. A small fraction of the original parasite population might always survive to the drug but it will be eventually removed by the immune system. However, the infection will not be cleared if the surviving fraction is too large due either to reduced sensitivity or to subcritical drug concentrations. Selection of resistant strains could occur when a particular drug is misused (Wernsdorfer 1991). The transmission of such parasites might also be enhanced by an increased production of gametocytes (Robert et al. 1996).

Humans and drugs - Readily absorbed drugs with a long half-life, like mefloquine and sulfadoxine-pyrimethamine (SP), can permit effective single dose treatment of malaria and the following chemoprophylactic period prevents infection for several weeks and may be important in recovery from anaemia. However, these drugs are likely to exert undesirable drug pressure for a long time once their concentrations drop below the critical threshold and may select resistant parasites. This has been shown in Kenya where a potent selective pressure for resistance operates even under conditions of supervised drug administration and optimal dosage (Watkins et al. 1997).

Vector and parasite - Vectors may be more receptive to resistant strains and may produce more parasites compared to sensitive strains. Enhanced drug pressure and uninhibited transmission might produce a fast selection and spread of resistant parasites (Wernsdorfer 1991).

NATIONAL DRUG POLICIES

It is the responsibility of National Programmes to develop a policy for malaria disease management as early diagnosis and adequate treatment remain the basic elements of any malaria control action

Fax: +32-3-247.6362. E-mail: UDAlessandro@proto.itg.be
Received 15 June 1998
Accepted 30 July 1998

(WHO 1993). The basis for the national drug policy should be a set of defined criteria for efficacy, side effects, costs and compliance. Rational prescribing in the public sector should be promoted through the development and introduction of treatment guidelines. These activities should be supported by the supply system to ensure that the drugs health workers have been trained to prescribe are actually available (Baudon 1995). Many of the factors influencing the national drug policy, such as parasite drug susceptibility, drug pricing and availability, will fluctuate over time and national guidelines for malaria treatment will require periodic re-assessment and revision (WHO 1996). This should be done on the basis of reliable information that, in case of parasite drug susceptibility should be collected by means of an appropriate surveillance system.

Implementing a change in drug policy is a major undertaking that can take several years before being fully operational. The 'reaction-time' could be quite long in countries where the capabilities of training and re-training of health workers is limited and the distribution system of drugs is complex, fragmentary and uncontrolled. A study done in Congo in 1993 has shown that 77% of General Practitioners (GPs) did not know the national drug policy and that they continued to apply a strategy (weekly malaria chemoprophylaxis in children under 5) abandoned seven years before (Baudon 1995). In optimal conditions the reaction time has been estimated to at least two years (Baudon 1995). Considering these difficulties, once the decision of changing the first line drug has been taken there is no go back.

SURVEILLANCE

One of the major questions is whether the first-line drug is still 'working'. Unfortunately, the criteria on which this decision can be taken are not clear (Bloland et al. 1993). The standard methods to assess the efficacy of a given drug can be divided in two broad groups: *in vivo* and *in vitro* tests.

In vivo test - During the *in vivo* test the recommended dose of an antimalarial drug is administered to infected subjects and the parasite's response in the host is assessed. The test could be done on symptomatic or asymptomatic people (Wernsdorfer & Payne 1988) with a 7-day follow-up (Prasad et al. 1990). It assesses only the initial parasitological response and, to a limited extent, clinical response to therapy during the follow-up. However, it does not address the implications and manifestations of persistent parasitaemia occurring after poor response, for example its impact on other condition such as anemia and malnutrition (Bloland et al. 1993). A study carried out in Malawi and in Kenya among young children with clinical malaria showed a shortening of the duration of clinical improvement and a decreased haematological recovery after therapy with chloroquine as compared to SP. In addition the health care system is burdened by both the drug cost and the manpower requirements of frequent and repetitive visits for treatment with a poorly efficacious drug. The new 14-days *in vivo* test proposed by WHO (WHO 1996) tries to address this problem by looking mainly at treatment failure and change in the haematological parameters (Hb/PCV at day 0 and 14). It is a more clinical test, carried out on sick children (fever + parasitaemia) aged 6 months-5 years (D'Alessandro et al. 1997). This test does not permit a quantitative assessment of the drug sensitivity of individual parasite populations, may occasionally be influenced by the abnormal fate of the drug in individual patients and it is influenced by the immunological host response to the parasite. However, it gives an information that is closer to real-life situation and therefore essential in deciding drug policy changes.

In vitro test - The *in vitro* test consists in measuring the inhibition of schizont maturation by increasing doses of a given antimalarial drug. It allows the quantitative measurement of drug response, permits to test several drugs at once and imitate the non-immune state. However, it is generally held that the *in vitro* tests do not reflect the degree of *in vivo* resistance, since the latter is substantially determined by factors related to the host's response (Draper et al. 1988). Furthermore, it is estimated that the technical capacity to conduct the necessary assays is difficult to develop and maintain in national malaria control programmes. Therefore, *in vitro* testing cannot substitute *in vivo* observations of malaria therapy and is inappropriate for making policy decision on drug use. Nevertheless, *in vitro* tests may provide an early warning of impeding resistance before this becomes clinically apparent. The optimal deployment of *in vitro* tests should be to define specific issues related to temporal and geographical trends of parasite's response to drugs. Such issues include the longitudinal follow-up of parasite drug susceptibility, monitoring the patterns of parasite cross-resistance to different drugs and the establishment of baseline data on the susceptibility of local parasites to new drugs (WHO 1990).

MOLECULAR METHODS FOR SURVEILLANCE

Recently, molecular diagnostic methods for detecting resistant parasites have been proposed for monitoring the level and spread of resistance (Plowe et al. 1995). The methods are suited for use on large numbers of samples in a laboratory in a malaria-endemic country and have major advan-

tages on *in vitro* tests that require parasite cultivation and take days to perform (Plowe et al. 1996). These molecular tools are based on the detection by PCR of point mutations in the parasite genes responsible for *in vitro* resistance. Presently, several mutations associated with resistance to SP have been identified while the role of those linked to chloroquine resistance is more controversial. The latter has been associated to *Pfmdr1* gene polymorphysm, although multiple mutations in different genes are probably required for clinical resistance (Foote et al. 1990). Wellems et al. (1990) have linked chloroquine resistance to a single genetic locus yet to be identified. The point mutations linked to SP resistance have been observed in the parasite genes encoding for dihydrofolate reductase (DHFR) and dihydropteroate synthetase (DHPS), the targets for pyrimethamine and sulfadoxine respectively (Wang et al. 1995). A serine in position 108 of the DHFR gene is linked to *in vitro* sensitivity to both pyrimethamine and cycloguanil. A mutation to asparigine at position 108 seems to be the key mutation for conferring *in vitro* pyrimethamine resistance (de Pecoulas et al. 1996), although a genotype without this mutation has been recently described (Wang et al. 1997). An asparagine to isoleucine change at position 51 and a cysteine to arginine at position 59 appear to modulate higher levels of *in vitro* pyrimethamine resistance when they occur with the asparagine-108 mutation, and an isoleucine to leucine mutation at position 164 in combination with the asparagine-108 and arginine-59 mutations has been found in *P. falciparum* lines that are highly resistant to both pyrimethamine and cycloguanil (Basco et al. 1995, Reeder et al. 1996). Point mutations of the DHPS gene have been less extensively studied. Thirteen variants over the wild type have been identified so far in samples from different countries, the most common alteration being in position 437 (Wang et al. 1997). A certain amount of correlation was found between the prevalence of known DHFR and DHPS mutations and increasing levels of *in vivo* SP resistance in four different countries: Mali, Kenya, Malawi, Bolivia (Plowe et al. 1997). Different degrees of *in vivo* parasitological resistance and clinical failure to SP might be due to the progressive accumulation of DHFR and DHPS mutations (Plowe et al. 1997). However, this still needs to be established by carrying out prospective studies on the relationship between parasite genotype and clinical outcome in individual infections treated with SP. No relation between clinical SP resistance and mutations in the DHFR gene could be established in a study carried out in Tanzania. Although all the isolates showed a point mutation in at least one codon of the DHFR gene, only 43% of the children had detectable parasitaemia seven days after treatment (Jelinek et al. 1997).

The characterisation of the two polymorphic merozoite surface antigens, MSP1 and MSP2 has also been used to establish whether a parasitaemia observed after treatment is caused by a recrudescence of drug-resistant parasites or by a new infection (Babiker et al. 1994, Al-Yaman et al. 1997). This could be an important information when carrying out *in vivo* tests, particularly in areas with a considerable amount of transmission where, after a certain time, it is impossible to distinguish between recrudescence and new infection. However, it is unlikely that national malaria control programmes will routinely use such methods.

In vitro test results have been used as the golden standard to establish the link between resistance to SP and point mutations in the DHFR and DHPS genes. Therefore, molecular tests identifying mutant parasites might replace *in vitro* tests, as they are easier to perform and can be carried out on a larger number of samples. They will complement the information on treatment failure and might actually identify areas where a given drug is likely to be less efficacious. However, for antimalarial drugs other than SP, these tests are not available yet and drug susceptibility of different parasite populations will continue to be established by *in vitro* tests. Decisions on the national drug policy will continue to be based on the results of *in vivo* tests as these reflect more closely the therapeutic efficacy of a given drug.

REFERENCES

Al-Yaman F, Genton B, Reeder JC, Anders RF, Alpers MP 1997. Evidence that recurrent *Plasmodium falciparum* infection is caused by recrudescence of resistant parasites. *Am J Trop Med Hyg 56*: 436-439.

Babiker H, Randford-Cartwright L, Sultan A, Satti G, Walliker D 1994. Genetic evidence that RI choloroquine resistance of *Plasmodium falciparum* is caused by recrudescence of resistant parasites. *Trans R Soc Trop Med Hyg 88*: 328-331.

Basco LK, de Pecoulas PE, Wilson CM, Le Bras J, Mazabraud A 1995. Point mutations in the dihydrofolate reductase-thymidylate synthase gene and pyrimethamine and cycloguanil resistance in *Plasmodium falciparum*. *Mol Biochem Parasitol 69*: 135-138.

Baudon D 1995. La prise de décision dans la politique d'utilisation des antipaludiques en réponse à une modification de l'efficacité de la chloroquine. Réflexions appliquées à l'Afrique. *Méd Trop 55*: 37S-40S.

Bloland PB, Lackritz EM, Kazembe PN, Were JBO, Steketee R, Campbell CC 1993. Beyond chloroquine: implications of drug resistance for evaluating malaria therapy efficacy and treatment policy in Af-

rica. *J Infect Dis 167*: 932-937.

D'Alessandro U, Mutabingwa TK, Ouma JH, Watkins WM 1997. Report of the workshop on monitoring antimalarial treatment efficacy: an East African regional network. Malaria Consortium, UK.

De Pecoulas PE, Basco LK, Le Bras J, Mazabraud A 1996. Association between antifol resistance *in vitro* and DHFR gene point mutation in *Plasmodium falciparum* isolates. *Trans R Soc Trop Med Hyg 90*: 181-182.

Draper CC, Hills M, Kilimali VAEB, Brubaker G 1988. Serial studies on the evolution of drug resistance in malaria in an area of East Africa: findings from 1979 up to 1986. *J Trop Med Hyg 91*: 265-273.

Foote SJ, Kyle DE, Martin RK, Oduola AM, Forsyth K, Kemp DJ, Cowman AF 1990. Several alleles of the multidrug-resistance gene are closely linked to chloroquine resistance in *Plasmodium falciparum*. *Nature 345*: 255-258.

Jelinek T, Ronn AM, Curtis J, Duraisingh MT, Lemnge MM, Minha J, Bygbjerg IC, Warhurst DC 1997. High prevalence of mutations in the dihydrofolate reductase gene of *Plasmodium falciparum* in isolates from Tanzania without evidence of an association to clinical sulfadoxine/pyrimethamine resistance. *Trop Med Intern Hlth 2*:1075-1079.

Plowe CV, Djimde A, Bouare M, Doumbo O, Wellens TE 1995. Pyrimethamine and proguanil resistance-conferring mutations in *Plasmodium falciparum* dihydrofolate reductase: polymerase chain reaction methods for surveillance in Africa. *Am J Trop Med Hyg 52*: 565-568.

Plowe CV Djimde A, Wellems TE, Diop S, Kouriba B, Doumbo OK 1996. Community pyrimethamine-sulfadoxine use and prevalence of resistant *Plasmodium falciparum* genotypes in Mali: a model for deterring resistance. *Am J Trop Med Hyg 55*: 467-471.

Plowe CV, Cortese JF, Djimde A, Nwanyanwu OC, Watkins WM, Winstanley PA, Estrada-Franco JG, Mollinedo RE, Avila JC, Cespedes JL, Carter D, Doumbo OK 1997. Mutations in *Plasmodium falciparum* dihydrofolate reductase and dihydropteroate synthase and epidemiological patterns of pyrimethamine-sulfadoxine use and resistance. *J Infect Dis 176*: 1590-1596.

Prasad RN, Prasad H, Virk KJ, Sharma VP 1990. Application of a simplified *in-vivo* test system for determining chloroquine resistance in *Plasmodium falciparum*. *Bulletin WHO 68*: 755-758.

Reeder JC, Rieckmann KH, Genton B, Lorry K, Wines B, Cowman AF 1996. Point mutations in the dihydrofolate reductase and dihydropteroate synthetase genes and *in vitro* susceptibility to pyrimethamine and cycloguanil of *Plasmodium falciparum* isolates from Papua New Guinea. *Am J Trop Med Hyg 55*: 209-213.

Robert V, Molez JF, Trape JF 1996. Short report: gametocytes, chloroquine pressure, and the relative parasite survival advantage of resistant strains of falciparum malaria in West Africa. *Amer J Trop Med Hyg 55*: 350-351.

Wang P, Brooks DR, Sims PFG, Hyde JE 1995. A mutation-specific PCR system to detect sequence variation in the dihydropteroate synthetase gene of *Plasmodium falciparum*. *Mol Biochem Parasitol 71*: 115-125.

Wang P, Lee CS, Bayoumi R, Djimde A, Doumbo O, Swedberg G, Dao LD, Mshinda H, Tanner M, Watkins WM, Sims PFG, Hyde JE 1997. Resistance to antifolates in *Plasmodium falciparum* monitored by sequence analysis of dihydropteroate synthetase and dihydrofolate reductase alleles in a large number of field samples of diverse origins. *Mol Biochem Parasitol 89*: 161-177.

Watkins WM, Mberu EK, Winstanley PA, Plowe CV 1997. The efficacy of antifolate antimalarial combinations in Africa: a predictive model based on pharmacodynamic and pharmacokinetic analyses. *Parasitol Today 13*: 459-464.

Wellems TE, Panton LJ, Gluzman IY, do Rosario VE, Gwadz RW, Walker-Jonah A, Krogstad DJ 1990. Chloroquine resistance not linked to *mdr*-like genes in a *Plasmodium falciparum* cross. *Nature 345*: 253-255.

Wernsdorfer WH, Payne D 1988. Drug sensitivity tests in malaria parasites, p. 1765-1794. In WH Wernsdorfer, I McGregor (eds), *Malaria - Principles and Practice of Malariology*, Churchill Livingstone, Edinbourg, London, Melbourn, New York.

Wernsdorfer WH 1991. The development and spread of drug-resistant malaria. *Parasitol Today 11*: 297-303.

WHO - World Health Organization 1990. Practical chemotherapy of malaria. Technical Report Series 805. WHO, Geneva.

WHO - World Health Organization 1993. Implementation of the global malaria control strategy. Report of a WHO Study Group on the implementation of the Global Plan of Action for Malaria Control 1993-2000. WHO, Geneva.

WHO - World Health Organization 1996. Assessment of therapeutic efficacy of antimalarial drugs for uncomplicated *falciparum* malaria in areas with intense transmission. WHO/MAL/96. 1077.

Allelic Diversity at the *Merozoite Surface Protein-1* (MSP-1) Locus in Natural *Plasmodium falciparum* Populations: a Brief Overview

Marcelo U Ferreira/+, Osamu Kaneko*, Masatsugu Kimura, Qing Liu***, Fumihiko Kawamoto***, Kazuyuki Tanabe******

Departamento de Parasitologia, ICB, Universidade de São Paulo, Av. Prof. Lineu Prestes 1374, 05508-900 São Paulo, SP, Brasil *Department of Medical Zoology **Laboratory of Biophysics, Osaka City University Medical School, Osaka, Japan ***Department of International Health, Nagoya University School of Medicine, Nagoya, Japan ****Laboratory of Biology, Osaka Institute of Technology, Osaka, Japan

The merozoite surface protein-1 (MSP-1) *locus of* Plasmodium falciparum *codes for a major asexual blood-stage antigen currently proposed as a major malaria vaccine candidate. The protein, however, shows extensive polymorphism, which may compromise its use in sub-unit vaccines. Here we compare the patterns of allelic diversity at the* MSP-1 *locus in wild isolates from three epidemiologically distinct malaria-endemic areas: the hypoendemic southwestern Brazilian Amazon (n = 54), the mesoendemic southern Vietnam (n = 238) and the holoendemic northern Tanzania (n = 79). Fragments of the variable blocks 2, 4a, 4b and 6 or 10 of this single-copy gene were amplified by the polymerase chain reaction, and 24* MSP-1 *gene types were defined as unique combinations of allelic types in each variable block. Ten different* MSP-1 *types were identified in Brazil, 23 in Vietnam and 13 in Tanzania. The proportion of genetically mixed infections (isolates with parasites carrying more than one* MSP-1 *version) ranged from 39% in Brazil to 44% in Vietnam and 60% in Tanzania. The vast majority (90%) of the typed parasite populations from Brazil and Tanzania belonged to the same seven most frequent* MSP-1 *gene types. In contrast, these seven gene types corresponded to only 61% of the typed parasite populations from Vietnam. Non-random associations were found between allelic types in blocks 4a and 6 among Vietnamese isolates, the same pattern being observed in independent studies performed in 1994, 1995 and 1996. These results suggest that* MSP-1 *is under selective pressure in the local parasite population. Nevertheless, the finding that similar* MSP-1 *type frequencies were found in 1994 and 1996 argues against the prominence of short-term frequency-dependent immune selection of* MSP-1 *polymorphisms. Non-random associations between* MSP-1 *allelic types, however, were not detected among isolates from Brazil and Tanzania. A preliminary analysis of the distribution of* MSP-1 *gene types per host among isolates from Tanzania, but not among those from Brazil and Vietnam, shows significant deviation from that expected under the null hypothesis of independent distribution of parasites carrying different gene types in the human hosts. Some epidemiological consequences of these findings are discussed.*

Key words: *Plasmodium falciparum* - malaria - allelic diversity - merozoite surface protein-1 - population genetics - vaccine candidate

The polymorphic *merozoite surface protein-1 (MSP-1)* of *Plasmodium falciparum* is a major asexual blood-stage malaria vaccine candidate (Holder 1996). Comparisons of nucleotide sequences led to the identification of seven variable blocks in the gene, which are interspersed with five conserved and five semi-conserved blocks (Fig. 1). There are essentially two versions of each block, named after the representative isolates MAD20 and K1 (Tanabe et al. 1987). A major exception to this dimorphic rule is the variable block 2, that has a third version originally described in the isolate RO33 (Certa et al. 1987). Most allelic diversity is generated by intragenic recombination between

Supported by grants from the Ministry of Education, Science, Sports and Culture of Japan, Toyota Foundation, the Program for Malaria Control in the Amazon of the National Health Foundation (Brazilian Ministry of Health), Fundação de Amparo à Pesquisa do Estado de São Paulo, Brazil, and the UNDP/World Bank/World Health Organization Special Programme for Research and Training in Tropical Diseases. MUF was supported by a research student scholarship from the Ministry of Education, Science, Sports and Culture of Japan.
+Corresponding author: Fax: +55-11-818.7417. E-mail: muferrei@usp.br
Received 15 June 1998
Accepted 30 July 1998

these representative sequences at the 5' end of the gene, within blocks 3, 4 and 5. Minor differences also exist between homologous versions of the same variable block, and nucleotide substitutions (most of which are dimorphic) occur in semi-conserved and conserved blocks (Tanabe et al. 1987).

Major *MSP-1* gene types may be defined as unique combinations of: (a) one of three versions of block 2 (MAD20, K1 or RO33), (b) one of four possible versions of block 4, because recombination within this region generates MAD20/K1 and K1/MAD20 hybrids in addition to the 'pure' allelic types MAD20 and K1 (Conway et al. 1991b, Kaneko et al. 1996), and (c) one of two versions (MAD20 or K1) of the segment between the variable blocks 6 and 16, that comprises about 60% of the gene. Recombination events have not been described in this portion of the gene (Tanabe et al. 1987, 1989, Peterson et al. 1988, Conway et al. 1991b, Jongwutiwes et al. 1991, Kaneko et al. 1996, 1997). Therefore, the 24 *MSP-1* gene types shown in Table I may theoretically be observed in natural parasite populations (Kaneko et al. 1997).

The extent of allelic diversity in different malaria-endemic areas should be evaluated if the variable domains of *MSP-1*, that are highly immunogenic (Holder & Riley 1996), are to be included in subunit malaria vaccines. A novel polymerase chain reaction (PCR)-based strategy was recently developed to group clinical isolates of *P. falciparum* into the 24 *MSP-1* gene types defined in Table I (Kaneko et al. 1997). This strategy has been successfully applied to type wild isolates from the mesoendemic southern Vietnam (Kaneko et al. 1997, Ferreira et al. 1998b), the hypoendemic Brazilian Amazon (Ferreira et al. 1998a), and the holoendemic Tanzania (Ferreira et al. 1998c). In this communication we analyze available data regarding complete *MSP-1* typing of isolates from these three malaria-endemic areas.

TABLE I

Merozoite surface protein-1 (MSP-1) gene types defined as unique combinations of allelic types in each variable block

Gene type[a]	Variable block 2	4a	4b	10
1	K1	K1	K1	K1
2	MAD20	K1	K1	K1
3	R033	K1	K1	K1
4	K1	MAD20	K1	K1
5	MAD20	MAD20	K1	K1
6	RO33	MAD20	K1	K1
7	K1	K1	MAD20	K1
8	MAD20	K1	MAD20	K1
9	RO33	K1	MAD20	K1
10	K1	MAD20	MAD20	K1
11	MAD20	MAD20	MAD20	K1
12	RO33	MAD20	MAD20	K1
13	K1	K1	K1	MAD20
14	MAD20	K1	K1	MAD20
15	RO33	K1	K1	MAD20
16	K1	MAD20	K1	MAD20
17	MAD20	MAD20	K1	MAD20
18	RO33	MAD20	K1	MAD20
19	K1	K1	MAD20	MAD20
20	MAD20	K1	MAD20	MAD20
21	RO33	K1	MAD20	MAD20
22	K1	MAD20	MAD20	MAD20
23	MAD20	MAD20	MAD20	MAD20
24	RO33	MAD20	MAD20	MAD20

a: each gene type is defined as a unique combination of allelic types detected in the variable blocks 2, 4a (5' segment of block 4), 4b (3' segment of block 4) and 6-16 of the *MSP-1* gene. Since there is no recombination at the central and C-terminal portions of this gene, the allelic type detected in block 10 is considered to be the same for the variable blocks 6, 8, 14 and 16. Allelic types are named after the reference isolates MAD20, K1 and RO33.

MATERIALS AND METHODS

Table II summarizes basic information regarding typed *P. falciparum* isolates in each malaria-endemic area. Genomic DNA was extracted directly from the blood of *P. falciparum*-infected patients, without previous *in vitro* cultivation of parasites. Locations of the oligonucleotide primers are shown in Fig. 1. Primer sequences and PCR protocols are given elsewhere (Kaneko et al. 1997). The basic PCR-based typing procedure developed by Kaneko et al. (1997) may be described as it follows:

First step - Block 2 was typed in three separate reactions with the allelic specific forward primers M2F, K2F and R2F and the common reverse primer C3R.

Second step - The gene fragments between the conserved block 5 and the variable block 6 were amplified in two separate reactions with the common forward primer C5F and the type-specific reverse primers M6R or K6R. Alternatively, block 10 was typed with the semi-conserved forward primer C9F and the type-specific reverse primers M10R and K10R. Since there is no recombination between blocks 6 and 16, the allelic type found in blocks 6 or 10 is the same for variable blocks 8, 14 and 16.

Third step - Segments between blocks 2 and 6 were amplified in three separate reactions with the type-specific forward primers M2F, K2F or R2F, and the type-specific reverse primers M6R or K6R. The PCR fragments amplified in this step were used as template in the next step. As an alternative, this

TABLE II

Recent polymerase chain reaction-based studies involving complete typing of the *merozoite surface protein-1* gene in natural *Plasmodium falciparum* populations

Area	Malaria endemicity	No. of typed isolates	Reference
Brazilian Amazon	Low	54	Ferreira et al. 1998a
Southern Vietnam	Intermediate	136	Kaneko et al. 1997
Southern Vietnam	Intermediate	102	Ferreira et al. 1998b
Northern Tanzania	High	79	Ferreira et al. 1998c

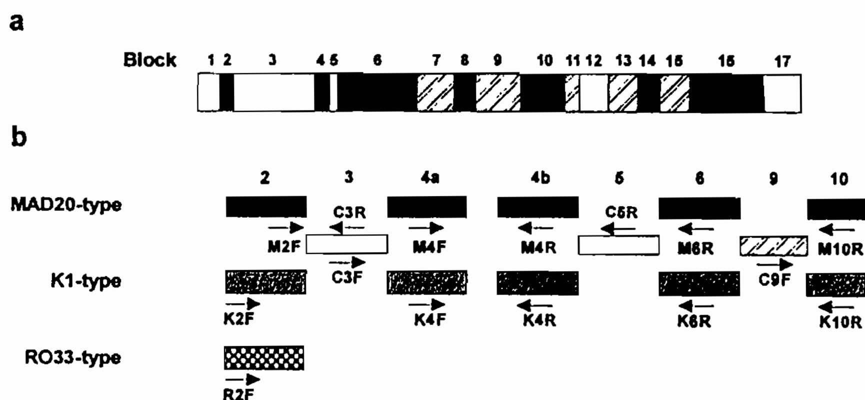


Fig. 1-a: structure of the *merozoite surface protein-1* gene of *Plasmodium falciparum*. Conserved, semi-conserved and variable blocks of the gene are shown as open, hatched and closed boxes, respectively. Block numbers are after Tanabe et al. (1987); b: locations and directions of the oligonucleotide primers used to type blocks 2, 4a, 4b, 6, and 10 are also indicated. Redrawn from Kaneko et al. 1997.

template may be prepared with the conserved forward primer C3F and the conserved reverse primer C5R.

Forth step - Block 4 was typed by nested PCR in four separate reactions with the type-specific forward primers M4F or K4F and type-specific reverse primers M4R or K4R. As an alternative, the first step may be eliminated, and block 2 may be typed by detecting allelic-specific fragments in the second step (Ferreira et al. 1998b).

The detection of PCR products in the expected size ranges after 1.5-2% agarose gel electrophoresis defined the presence of each allelic type in blocks 2, 6 or 10, 4a and 4b. As *MSP-1* is a single-copy gene in the haploid genome of blood-stage parasites, we consider that isolates harboring more than one gene type have mixed infections with genetically distinct *P. falciparum* subpopulations. Each subpopulation may be separately typed by this approach (Kaneko et al. 1997).

RESULTS AND DISCUSSION

Are all theoretically possible MSP-1 *gene types found in natural* Plasmodium falciparum *populations?* - As shown in Fig. 2, all but one of the 24 possible *MSP-1* gene types were detected in mesoendemic Vietnam. Only gene type 11 was absent in that area at both occasions (Kaneko et al. 1997, Ferreira et al. 1998b). This suggests that almost all possible combinations of *MSP-1* allelic types may be found in parasites that are able to infect human hosts. In contrast, only 10 and 13 *MSP-1* types were found in hypoendemic Brazil and holoendemic Tanzania, respectively. Moreover, essentially the same *MSP-1* gene types were found to predominate in both countries, and about 90% of the typed parasite populations belonged to the seven most common gene types, namely the types 13, 16, 17, 18, 22, 23 and 24 as defined in Table I (Ferreira et al. 1998c). Nevertheless, these seven gene types were found in only 61% of the typed parasite populations in Vietnam.

Is there any association between the extent of MSP-1 *diversity and the intensity of malaria transmission?* - If we compare the proportions of genetically mixed infections (that is, patients harboring more than one *MSP-1* gene type) and the average number of *MSP-1* gene types found per patient, an apparent positive association is found between malaria endemicity and *MSP-1* diversity (Table III). However, if we compare the number of different *MSP-1* gene types found in each en-

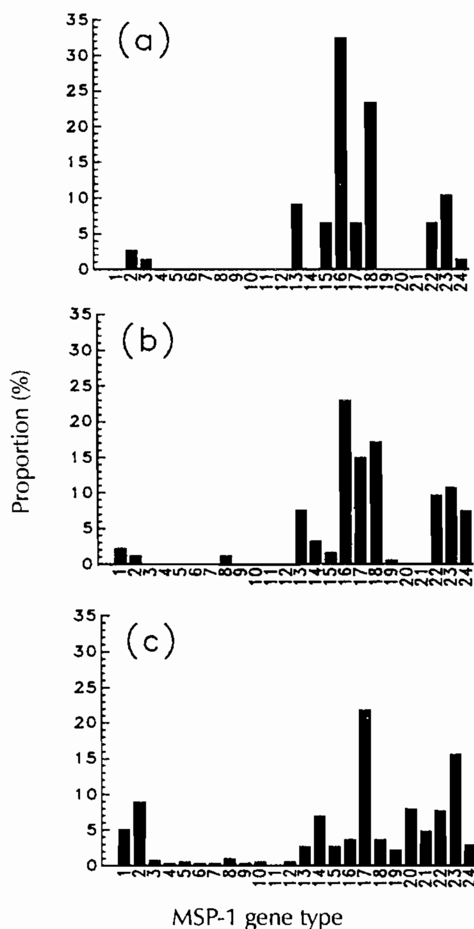


Fig. 2-a: frequency distribution of the *merozoite surface protein-1 (MSP-1)* gene types in 54 *Plasmodium falciparum* isolates collected in July 1995 in the city of Porto Velho, State of Rondônia, southwestern Brazilian Amazon (Ferreira et al. 1998a); b: frequency distribution of the *MSP-1* gene types in 79 *P. falciparum* isolates collected between July and September 1996 in the city of Tanga and the nearby village of Pangani, in northern Tanzania (Ferreira et al. 1998c); c: frequency distribution of the *MSP-1* gene types in 238 *P. falciparum* isolates collected between July 1994 and July 1996 from malaria patients belonging to the ethnic majority Kinh and the minority K'ho living in the towns of Bao Loc and Phu Rieng and nearby areas in southern Vietnam (Kaneko et al. 1997, Ferreira et al. 1998b). The 24 *MSP-1* gene types are numbered as in Table I.

demic area, no such association can be detected. Therefore, despite the fact that most infected hosts in Tanzania carry two or more parasite clones which may be ingested by the vector and recombine during meiosis, the resulting repertoire of MSP-1 variants seems to be relatively restricted in human hosts, if compared with the situation found in Vietnam. Strong selective pressure related to the sequential use of several different antimalarials in a few years, in the context of multi-drug resistance, may have resulted in increased genetic diversity of *P. falciparum* populations in Vietnam.

Are the patterns of MSP-1 *diversity temporally stable in a given malaria-endemic area?* - Fig. 3 compares the distribution of *MSP-1* gene types in parasite populations sampled in the same communities in southern Vietnam at intervals of 12 months (Fig. 3a) and 18-24 months (Fig. 3b). There is no significant difference when both pairs of frequency distributions are compared. The stability in the frequencies of *MSP-1* gene types over periods of 12-24 months does not imply that long-term changes can be ruled out. Under the present conditions of malaria transmission in southern Vietnam, just a few infections with parasites carrying distinct versions of the MSP-1 antigen are expected per host at a one-year or two-year interval. As a consequence, natural acquisition of effective anti-MSP-1 immunity may occur at a rather slow rate. We have now examined this issue in relation to the Brazilian Amazon by typing *MSP-1* variable blocks in *P. falciparum* isolates collected over a period of 12 years (LA Silveira & MU Ferreira, unpublished data).

Are there non-random associations between MSP-1 *variable blocks in natural parasite populations?* - If intragenic recombination occurs frequently at the *MSP-1* locus in the absence of major selective constraints, the distribution of *MSP-1* gene types would be described by a simple probability model analogous to those used in population genetics to estimate expected frequencies of multiple-locus genotypes (Tibayrenc 1995). For instance, the expected frequency of gene type 1 (as defined in Table I) is given by multiplying the

TABLE III

The extent of allelic diversity at the *merozoite surface protein-1 (MSP-1)* locus in natural *Plasmodium falciparum* populations from areas with different levels of malaria endemicity

Area	Malaria endemicity	No. of *MSP-1* gene types detected by PCR	Proportion (%) of isolates with > 1 *MSP-1* type	Average no. of *MSP-1* types per patient
Brazil	Low	10	39	1.42
Vietnam	Intermediate	23	44	1.76
Tanzania	High	13	60	2.37

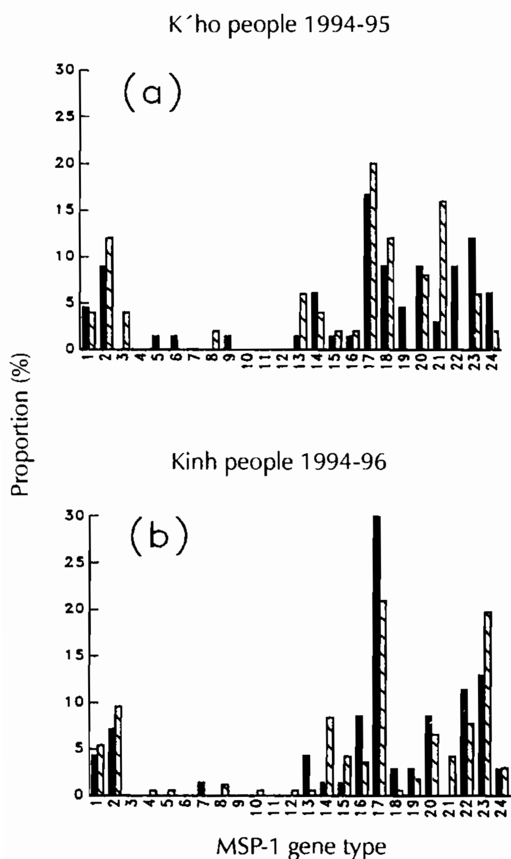


Fig. 3-a: frequency distribution of *merozoite surface protein-1(MSP-1)* gene types in isolates from K'ho people living in hill areas surrounding Bao Loc, southern Vietnam, collected between July-August 1994 (n = 34) (closed bars) and in August 1995 (n = 28) (striped bars) (redrawn from Kaneko et al. 1997); b: frequency distribution of *MSP-1* gene types in isolates from Kinh people living in the town of Bao Loc, southern Vietnam, collected between July-August 1994 (n = 44) (closed bars) and between January-July 1996 (n = 95) (striped bars) (redrawn from Ferreira et al. 1998b).

observed frequencies of the allelic type K1 in blocks 2, 4a, 4b and 6-16 in a given population. Fig. 4 shows expected frequencies of *MSP-1* gene types under the null hypothesis of random association of allelic types (MAD20, K1 or RO33) in each variable block in Brazil and Tanzania. No significant difference between expected and observed frequencies was detected by the χ^2 test for goodness of fit in both cases (Ferreira et al. 1998a, c). In contrast, significant differences between expected and observed frequencies of *MSP-1* gene types were found in two surveys in southern Vietnam (Fig. 5). Non-random associations were found to occur, in both cases, between blocks 4a and 6-16: *MSP-1* gene types with concordant allelic families (either MAD20 or K1) in blocks 4a and 6 or 10 were found more frequently than expected (Kaneko et al. 1997, Ferreira et al. 1998b). The reasons why similar results were not found in holoendemic Tanzania and hypoendemic Brazil remain to be elucidated.

Non-random associations between allelic types may result from: (a) geographic isolation leading to random genetic drift, (b) limited chances for intragenic recombination during meiosis in the mosquito vector due to the presence of few different *MSP-1* versions and the low prevalence of mixed infections in human hosts, and (c) biological constrains which bias for particular associations. As

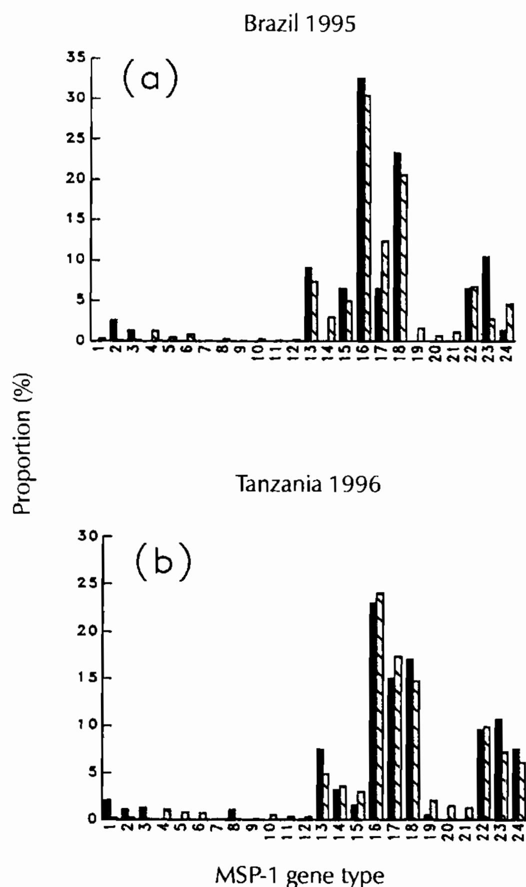


Fig. 4-a: expected (closed bars) and observed (striped bars) frequencies of *merozoite surface protein-1 (MSP-1)* gene types in Porto Velho, southwestern Brazilian Amazon (n = 54) (data from Ferreira et al. 1998a); b: expected (closed bars) and observed (striped bars) frequencies of *MSP-1* gene types in Tanga and Pangani, northern Tanzania (n = 79) (data from Ferreira et al. 1998c). Expected frequencies were computed under the null hypothesis of random associations of allelic types in variable blocks of the gene (see the text for details). There is no significant difference between expected and observed frequencies by χ^2 tests of goodness of fit.

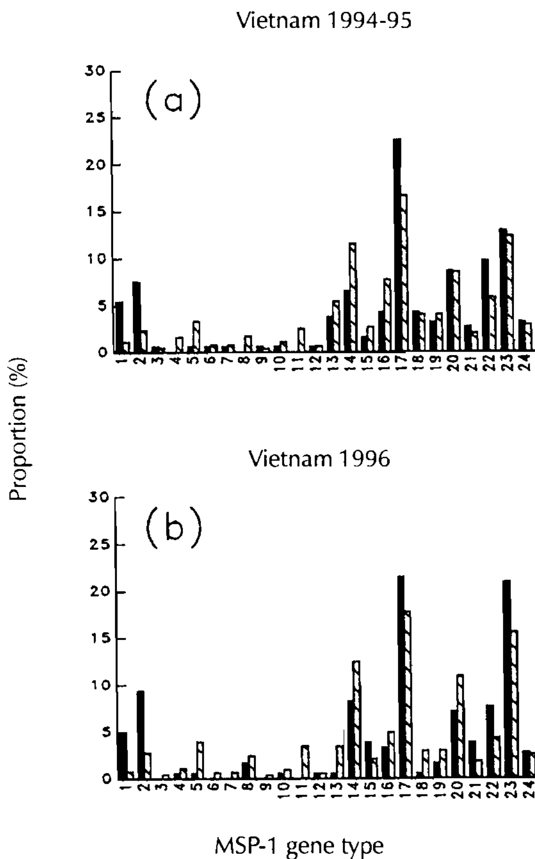


Fig. 5-a: expected (closed bars) and observed (striped bars) frequencies of *merozoite surface protein-1 (MSP-1)* gene types in isolates from Bao Loc, Vietnam, collected between July-August 1994 from both Kinh and K'ho people (n = 108) (data from Kaneko et al. 1997); b: expected (closed bars) and observed (striped bars) frequencies of *MSP-1* gene types in isolates from Bao Loc, Vietnam, collected between January-July 1996 from both K'ho and Kinh people (n = 102) (data from Ferreira et al. 1998b). Expected frequencies were computed under the null hypothesis of random associations of allelic types in variable blocks of the gene (see the text for details). In both cases significant differences between expected and observed frequencies were detected by χ^2 tests of goodness of fit.

discussed elsewhere, the first two possible explanations do not match available data from Vietnam, and speculations regarding the third hypothesis are limited by the fact that the function of *MSP-1* remains unknown (Ferreira et al. 1998b).

Are parasite populations carrying different MSP-1 *gene types independently distributed in the host population?* - A basic assumption of recent mathematical models of malaria transmission is that infections by different 'strains' are independent. This means that, in genetically mixed infections, a patient infected by a parasite carrying a given *MSP-1* gene type (for instance type 1) is as likely to be co-infected with a given second type (for instance type 2) as someone infected with any other *MSP-1* type. This does not take into account the possibilities of: (a) frequent multiple-clone infections by vectors carrying two or more gene types including recombinant gene types resulting from the union of two different clones from one previous host (Hill & Babiker 1995), and (b) either facilitation or competition between parasites carrying different versions of a polymorphic antigen which co-infect the same host (Gilbert et al. 1998). We applied here a simple statistical analysis to test this assumption.

The expected distribution of *MSP-1* gene types per host under the hypothesis of independent distribution of *MSP-1* gene types may be described as the sum of N independent binomial distributions, where N is the number of different *MSP-1* gene types observed in host population. The variance of this summed binomial distribution, or expected variance σ^2, was calculated and compared to the observed variance s^2 using a χ^2 test as described (Lotz & Font 1991). The difference between the expected and observed variances was statistically significant in Tanzania (Table IV), suggesting that the *MSP-1* gene types are not independently distributed in the host population. Therefore, we tested 21 possible pairwise associations between gene types in 2 x 2 contingency tables using either standard χ^2 or Fisher's exact tests when appropriate, with the significance level adjusted for multiple

TABLE IV

Statistical comparison of expected (σ^2) and observed (s^2) variances of the distribution of *merozoite surface protein-1 (MSP-1)* gene types among human hosts living in areas with different levels of malaria endemicity

Area	Malaria endemicity	Expected variance (σ^2)	Observed variance (s^2)	χ^2 (d. f.)[a]	P
Brazil	Low	1.04	0.32	16.45 (53)	> 0.05
Vietnam	Intermediate	1.35	1.42	105.65 (101)	> 0.05
Tanzania	High	1.59	2.10	103.08 (78)	< 0.05

a: degrees of freedom.

comparisons with the Bonferroni's correction (Lord et al. 1997). At least 1 pair of *MSP-1* gene types (18 and 24) was found to be positively associated ($P = 0.0008$ by χ^2 test). With the Bonferroni's correction applied to these data, an association is statistically significant at the 5% level if $P < 0.0024$. This means that the genetically similar types 18 and 24, that differ only in the block 4b allelic type (Table I), tend to co-occur more frequently than expected under the null hypotheses that they are independently transmitted. Nevertheless, departures from the null hypothesis of independent transmission were not detected in areas of lower endemicity such as Brazil and Vietnam (Table IV).

Mathematical models have recently regarded malaria as a heterogeneous disease caused by several independently transmitted and antigenically distinct parasite subpopulations or 'strains' that do not interact within the human hosts and are able to elicit 'strain'-specific protective immunity. These models estimate the basic reproduction number R_0 of malaria, defined as the average number of secondary infections generated by one primary infection in a fully susceptible population, as a weighted average of R_0 values for each 'strain'. This estimate is substantially lower than R_0 values obtained by conventional methods, suggesting that malaria eradication in Africa may be quite feasible (Gupta et al. 1994). Nevertheless, the finding that genetically and antigenically distinct parasite populations are not independently distributed in the human hosts in areas of high endemicity, such as northern Tanzania (Ferreira et al. 1998c) and the Gambia (Conway et al. 1991a), implies R_0 values considerably higher than those provided by the weighted average approach (Lord et al. 1997).

In conclusion, this study provides examples of the use of simple molecular and statistical approaches to investigate the extent of antigenic diversity in malaria parasites and to test hypotheses regarding the patterns of transmission and interaction of genetically distinct parasite subpopulations in endemic areas.

ACKOWLEDGMENTS

To HV Thien (Lam Dong II Provincial Hospital, Bao Loc, Vietnam), BT Ndawi (Primary Health Care Institute, Iringa, Tanzania), M Zhou and S Isomura (Nagoya University School of Medicine, Nagoya, Japan), AM Katzin and EAS Kimura (Department of Parasitology, University of São Paulo, Brazil).

REFERENCES

Certa U, Rotmann D, Matile H, Reber-Liske RA 1987. A naturally occuring gene encoding the major surface antigen precursor p190 of *Plasmodium falciparum* lacks tripeptide repeats. *EMBO J 6*: 4137-4142.

Conway DJ, Greenwood BM, McBride JS 1991a. The epidemiology of multiple-clone *Plasmodium falciparum* infections in Gambian patients. *Parasitology 103:* 1-6.

Conway DJ, Rosário V, Oduola AMJ, Salako AL, Greenwood BM, McBride JS 1991b. *Plasmodium falciparum*: intragenic recombination and nonrandom associations between polymorphic domains of the precursor to the major surface antigens. *Exp Parasitol 73:* 469-480.

Ferreira MU, Liu Q, Kaneko O, Kimura M, Tanabe K, Kimura EAS, Katzin AM, Isomura S, Kawamoto F 1998a. Allelic diversity at the *merozoite surface protein-1* locus of *Plasmodium falciparum* in clinical isolates from the southwestern Brazilian Amazon. *Am J Trop Med Hyg*, in press.

Ferreira MU, Liu Q, Zhou M, Kaneko O, Kimura M, Thien HV, Isomura S, Tanabe K, Kawamoto F 1998b. Stable patterns of allelic diversity at the merozoite surface protein-1 locus of *Plasmodium falciparum* in clinical isolates from southern Vietnam. *J Euk Microbiol 45:* 131-136.

Ferreira MU, Liu Q, Kimura M, Ndawi BT, Tanabe K, Kawamoto F 1998c. Allelic diversity in the merozoite surface protein-1 and epidemiology of multiple-clone *Plasmodium falciparum* infections in northern Tanzania. *J Parasitol*, in press.

Gilbert SC, Plebanski M, Gupta S, Morris J, Cox M, Aidoo M, Kwiatkowski D, Greenwood BM, Whittle HC, Hill ASV 1998. Association of malaria parasite population structure, HLA, and immunological antagonism. *Science 279:* 1173-1177.

Gupta S, Trenholme K, Anderson RM, Day KP 1994. Antigenic diversity and transmission dynamics of *Plasmodium falciparum. Science 263:* 961-963.

Hill WG, Babiker HA 1995. Estimation of numbers of malaria clones in blood samples. *Proc R Soc London B 262:* 249-257.

Holder AA 1996. Preventing merozoite invasion of erythrocytes, p. 77-104. In SL Hoffman, *Malaria Vaccine Development. A Multi-immune Response Approach*, ASM Press, Washington D.C.

Holder AA, Riley EM 1996. Human immune response to *MSP-1*. *Parasitol Today 12:* 173-174.

Jongwutiwes S, Tanabe K, Nakazawa S, Uemura H, Kanbara H 1991. Coexistence of gp195 alleles of *Plasmodium falciparum* in a small endemic area. *Am J Trop Med Hyg 44:* 299-305.

Kaneko O, Jongwutiwes S, Kimura M, Kanbara H, Ishii A, Tanabe K 1996. *Plasmodium falciparum*: variation in block 4 of the precursor to the major surface proteins (MSP1) in natural populations. *Exp Parasitol 84:* 92-95.

Kaneko O, Kimura M, Kawamoto F, Ferreira MU, Tanabe K 1997. *Plasmodium falciparum*: allelic variation in the merozoite surface protein 1 in wild isolates from southern Vietnam. *Exp Parasitol 86:* 45-57.

Lord CC, Woolhouse MEJ, Barnard BJH 1997. Transmission and distribution of virus serotypes: African horse sickness in zebra. *Epidemiol Infect 118:* 43-50.

Lotz JM, Font WF 1991. The role of positive and negative interspecific associations in the organization of communities of intestinal helminths of bats. *Parasitology 103:*127-138.

Peterson MG, Coppel RL, Moloney MB, Kemp DJ 1988. Third form of the precursor to the major surface antigens of *Plasmodium falciparum. Mol Cell Biol 8:* 2664-2667.

Tanabe K, Murakami K, Doi S 1989. *Plasmodium falciparum*: dimorphism of the p190 alleles. *Exp Parasitol 68:* 470-403.

Tanabe K, Mackay M, Goman M, Scaiffe JG 1987. Allelic dimorphism in a surface antigen genc of the malaria parasite *Plasmodium falciparum. J Mol Biolol 195:* 273-287.

Tibayrenc M 1995. Population genetics of parasitic protozoa and other microorganisms. *Adv Parasitol 36:* 47-115.

Evaluation of DNA Recombinant Methodologies for the Diagnosis of *Plasmodium falciparum* and their Comparison with the Microscopy Assay

L Urdaneta/*, P Guevara/+, JL Ramirez*

Escuela de Malariología y Saneamiento Ambiental "Dr. Arnoldo Gabaldón", Maracay,Venezuela
*Grupo de Genética Molecular, Instituto de Biologia Experimental, Universidad Central de Venezuela, Caracas,Venezuela

Since 1984, DNA tests based on the highly repeated subtelomeric sequences of Plasmodium falciparum *(rep 20) have been frequently used in malaria diagnosis. Rep 20 is very specific for this parasite, and is made of 21 bp units, organized in repeated blocks with direct and inverted orientation. Based in this particular organization, we selected a unique consensus oligonucleotide (pf-21) to drive a PCR reaction coupled to hybridization to non-radioactive labeled probes. The pf-21 unique oligo PCR (pf-21-I) assay produced DNA amplification fingerprints when was applied on purified* P. falciparum *DNA samples (Brazil and Colombia), as well as in patient's blood samples from a large area of Venezuela. The performance of the Pf-21-I assay was compared against Giemsa stained thick blood smears from samples collected at a malaria endemic area of the Bolivar State, Venezuela, at the field station of Malariologia in Tumeremo. Coupled to non-radioactive hybridization the pf-21-I performed better than the traditional microscopic method with a r=1.7:1. In the case of mixed infections the r value of* P. falciparum *detection increased to 2.5:1. The increased diagnostic sensitivity of the test produced with this homologous oligonucleotide could provide an alternative to the epidemiological diagnosis of* P. falciparum *being currently used in Venezuela endemic areas, where low parasitemia levels and asymptomatic malaria are frequent. In addition, the DNA fingerprint could be tested in molecular population studies.*

Key words: *Plasmodium falciparum* - diagnosis - polymerase chain reaction - malaria

Plasmodium falciparum, the agent of the most lethal form of malaria, causes 1.5 to 2.7 million deaths each year, mostly among children. The incidence of malaria in the world is estimated to be 300-500 million clinical cases annually (WHO 1997). An approximate of 2,300 million people lived in areas with malaria risk, distributed in 100 endemic countries including Venezuela.

The Venezuela global incidence in 1997 was 28,056 malaria cases (Dirección de Endemias Rurales, Venezuelan Malaria Program, pers. commun.). Bolívar State located in the Amazon basin is responsible for 40-50% of the global incidence for the whole country. The number of malaria cases has been in a constant rise during the previous years, due mainly to local economic activities (gold and diamond mines), changes in human migration patterns and a high prevalence of *P. falciparum* resistant malaria to chloroquine and pyrimethamine/sulfadoxine drugs.

Studies on epidemiology of parasitic infections, on measures to control disease and clinical evaluation of treatments, all require identification of the infecting species. The identification of malaria parasite is usually performed with traditional microscopic diagnosis of Giemsa-stained thick blood, which is inexpensive and of easy application in the field. Nevertheless, the sensitivity of microscopic method depends on highly trained examiners, is time-consuming when large numbers of samples must be examined, and is thus not the most appropriate method for large-scale epidemiological surveys.

Assay strategies have been proposed that directly detect abundant parasite nucleic acid sequences, including repetitive DNA (Franzen et al. 1984, Barker et al. 1986, Oquendo et al. 1986, McLaughlin et al. 1987, Zolg et al. 1987) or ribosomal RNA (rRNA) (Lal et al. 1989, Wathers & McCutchan 1989).

Financial support from Escuela de Malariología y Saneamiento Ambiental "Dr. Arnoldo Gabaldón", Maracay-Venezuela, la Fundación para el Desarrollo de la Ciencia y la Tecnología de Aragua, Venezuela and Grupo de Genética Molecular del Instituto de Biología Experimental, Universidad Central de Venezuela, Venezuela.
+Corresponding author: Fax: +58-2-753-5897. E-mail: pguevara@fred.fhcrc.org
Received 15 June 1998
Accepted 30 July 1998

In the *P. falciparum* genome there are several repetitive sequences. One of these consists of 21-bp blocks, imperfectly repeated in tandem clusters oriented in opposite directions (rep-20). These rep-20 sequences are found in all chromosomes of the parasite (Oquendo et al. 1986). DNA probes directed to these repeats hybridize with *P. falciparum* strains from South America, Africa, and South and Southeast Asia (Buesing et al. 1987), and their sensitivity is comparable to that of conventional microscopy (Lanar et al. 1989). The sensitivity of DNA and RNA approaches have been further increased through the amplification of the target DNA using the polymerase chain reaction (PCR).

Considering the high specificity of the rep-20 for *P. falciparum* and its genomic organization, our objective was directed to design a diagnostic PCR assay coupled to non-radioactive hybridization, for the detection of *P. falciparum* in human blood samples. In the present communication we report this diagnostic assay using a single oligo and digoxigenin labeled probes detected by photoluminescence. Its application to blood samples obtained and processed in the endemic areas, compared to traditional thick smear technique is discussed. We also report the application of the assay to isolated *P. falciparum* DNA to produce amplification patterns differentiating strains of several geographical locations and proposing its use as a tool in molecular epidemiological studies.

MATERIALS AND METHODS

Blood samples collection and treatment - The sample blood was obtained from patients with malaria symptoms assisting to the malaria diagnostic post of the Venezuelan Malaria Program in Tumeremo, Bolívar State. The samples were collected during two visits of one week each to the endemic area during July 1993 and March 1994.

Duplicate samples of fingerprick blood were collected from 33 individuals in heparinized capillary tubes; 50 µl were deposited in Wathman paper filter, dried at room temperature and stored in individual and labeled sealed plastic bag, in case it was necessary to repeat the PCR assay. Another 50 µl were transferred to a 1.5 ml centrifuge tube in which the *P. falciparum* DNA was isolated using the chelex-100-iron protocol (Wooden et al. 1993).

Microscopic examination - Two thick blood films were prepared for each patient during the blood collection process. The blood smears were stained with Giemsa, and one reading was performed at field site under routine conditions of work (100 fields examined under oil immersion optics before a slide was considered negative), by an expert microscopist at the diagnostic post of the Malaria Program. For comparison purposes, a second microscopic diagnostic was done by one of us (L.U.); in this case, the sample was considered negative after 200 microscopic fields were examined. The parasite number was registered with respect to 200 white blood cells.

P. falciparum reference strain DNAs - *P. falciparum* reference strain DNAs were a gift of the following researchers: (1) one Colombian strain from Dr Moises Wasserman (Instituto Nacional de Higiene de Colombia), (2) 12 Brazilian isolates from Dr Hernan Del Portillo (Universidade de São Paulo, Brasil) and (3) 2 cultured strains from Dr Erlinda Sánchez (Universidad de Carabobo and Malariología, Venezuela). The parasite genomic DNA was extracted using proteinase K, followed by phenol-chloroform extraction and ethanol precipitation.

Primers selection and PCR assays - Two oligonucleotides were designed: (1) a primer of 16 bp, pf-16 (5'-ACT AAC TTA GGT CTTA-3'), and (2) a primer of 21bp pf-21 (5'-ATG TTA GTC AAC TTA AGA CCT-3'), both derived from the pf-21 consensus sequence reported by Oquendo et al. (1986).

The *P. falciparum* assay described by Tirasophon et al. (1991), was included in this study as a reference test. This PCR uses primers K114-P1 (5'-CGC TAC ATA TGC TAG TTG CCA GAC-3') and K114-P2 (5'-CGT GTA CCA TAC ATC CTA CCA AC-3') for amplification of a 206 bp fragment and, primer K1-14 (5'-GCT ATA ACC ACT ATT GCA ACG-3') for hybridization. We designated the PCR assays I to III according to the oligonucleotid used: PCR I: pf-21 primer alone, PCR II: two primers pf-16 and pf-21, and PCR III: reference test using oligos K114-P1 and K114-P2.

PCR reaction conditions - Amplification reaction mixtures consisted of 50 mM KCL, 10 mM Tris HCL pH 8.8, 1% Triton X-100, 3mM dNTPs, and 3 mM MgCl2. Primers pf-21 and pf-16 were at 4 µM. Primers K114-P1 and K114-P2 were at 1 µM. 1.25 units of Taq polymerase. The total volume was 25 µl. The specificity and sensitivity of the assays were tested against different *P. falciparum* DNA concentrations (200 ng/µl to 0.02 pg/µl), including in each case a negative control. Amplification products were analyzed by electrophoresis in 3% agarose gels and detected by staining with ethidium bromide.

PCR I and PCR II initial denaturation was done at 94ºC for 5 min followed by 35 cycles of 94ºC for 1 min, annealing at 40ºC for 1 min, and extension at 72ºC for 2 min, and a final elongation at 72ºC for 5 min.

PCR III initial denaturation at 94ºC for 5 min; 35 cycles of denaturation at 94ºC for 1 min, an-

nealing at 60ºC for 1 min, and extension at 72ºC for 1 min, and a final elongation at 72ºC for 5 min.

Blots and DNA hybridizations - Amplified DNAs were size fractionated by electrophoresis in 3% agarose gel and transferred onto nylon membranes (Hybond N+, Amersham) using the VacuGene apparatus (Pharmacia LKB) following the manufacturer instructions. Slot blots were performed as described in Davis et al. (1986).

The filters were pre-hybridized in 2X SSC, 10X Denhardt's solution (Maniatis et al. 1982) and 1.5% blocking reagent (Boehringer-Mannheim) for 1 hr at 37ºC. The hybridization with digoxigenin labeled oligonucleotides were incubated for 2 hr at 53ºC for pf-21, and 50ºC for K1-14. Filters were washed in 6X SSC-1% SDS at room temperature for 15 min, 2X SSC-0.1% at room temperature for 15 min, and finally 0.2X SSC-0.1% SDS at 37ºC for 15 min. In the case of DNA probes, the filters were pre-hybridized in 6X SSC, 5X Denhardt's, 0.5% SDS, 5% blocking reagent and 50% formamide for 2 hr at 37ºC. The hybridizations with digoxigenin labeled amplified products pf-21 and K114-P1/K114-P2 were incubated at 37ºC for 12-18 hr. Filters were washed every 15 min in 2X SSC-0.1% SDS at room temperature for 1 hr, 1X SSC-0.1% SDS at 68ºC each 15 min for 1 hr, and 0.1X SSC-0.1% SDS at 68ºC twice each 15 min. The maxim concentration used for both the oligonucleotides and the probes was 20 ng/μl of hybridization solution.

DNA probe labeling and detection - The Genius Kit from Boehringer-Mannheim was used for DNA labeling and detection by photoluminiscence. The oligonucleotides were labeled incorporating the dUTP-DGX using terminal transferase, following the manufacturer recommendations. The DNA probes were labeled by the multiprimer system incorporating the dUTP-DGX nucleotide. The photoluminescent substrate used was Lumi-phos 530 and the signals were detected using X-rays films.

RESULTS

PCR assays - Two PCR amplification systems were designed and standardized: PCR (I) pf-21 primer and PCR (II) pf-16+pf-21 primers. Both systems with target in the 21 bp repeat (rep 20). The reference system PCR (III) K114-P1/K114-P2 primers, was standardized to our conditions as described above. Fig. 1 shows the results of these amplification systems using DNA from two *P. falciparum* strains from Colombia. Both the PCR I and II directed to rep 20, yielded a broad range of amplified products, with band sizes from 200 bp to 3 Kb (Fig. 1A, C; lanes 1, 2), with higher intensity in the PCR II system. Similar results were reported by Barker et al. (1992), using complementary and degenerate primers directed against rep 20. Nevertheless, the use of two primers PCR II yielded nonspecific amplification with human DNA (Fig. 1C; lane 3) and negative control without DNA (Fig. 1C; lane 4). In the amplification with a single primer (PCR I) (Fig. 1A; lanes 1, 2), a broad range of reproducible amplified products was observed, clearly showing a 344 bp band for this *P. falciparum* isolate. No amplification was observed in the negative control (Fig. 1A; lanes 3, 4). The PCR III reference system yielded an expected single 206 bp band specific for *P. falciparum* (Fig. 1B; lanes 1, 2), without amplifying the human DNA (lane 3).

Considering the specificity and the amplification profile in *P. falciparum* DNA, the PCR I system was selected for detection of *P. falciparum* from field samples.

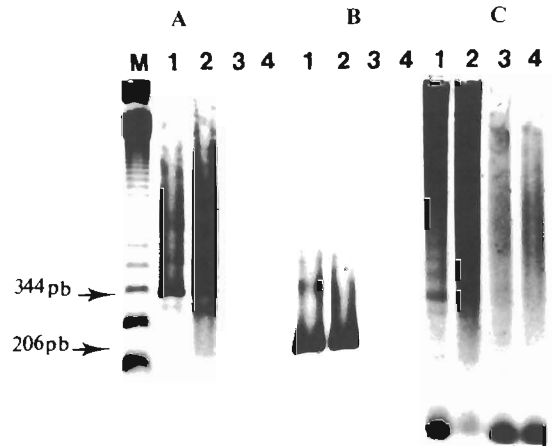


Fig. 1: PCR systems I, II and III. Electrophoresis in 3% agarose gel of amplified products by the systems. A: PCR I pf-21; B: PCR III K114-P1/K114-P2; C: PCR II pf-21+pf-16. Lanes - 1: 200 ng of *Plasmodium falciparum* purified DNA; 2: 50 ng of *P. falciparum* purified DNA; 3: negative control with human DNA; 4: negative control without DNA; M, 123 bp ladder.

Sensitivity of PCR I and PCR III systems - Once the specificity of these systems was determined, experiments were done to examine the detection sensitivity of both systems against serial dilutions of purified *P. falciparum* DNA. Fig. 2 shows the results of amplification products in 3% agarose gel electrophoresis for PCR I (Fig. 2A), PCR III (Fig. 2B) and Southern PCR III hybridizations (Fig. 2C). Both systems showed a detection level of 0.2 pg of *P. falciparum* DNA on agarose gels (Figs 2A, 2B; lane 7), equivalent to ten parasites. The South-

ern hybridization tests with digoxigenin labeled probe, yielded the following results: (a) the PCR I hybridized to the pf-21 primer showed a sensitivity level lower than the one observed in the electrophoresis, suggesting problems with the hybridization conditions (results not shown); (b) the PCR III hybridization (Fig. 2C) showed an increase in the sensitivity to 0.02 pg of DNA, equivalent to one parasite (Fig. 2C; lane 8).

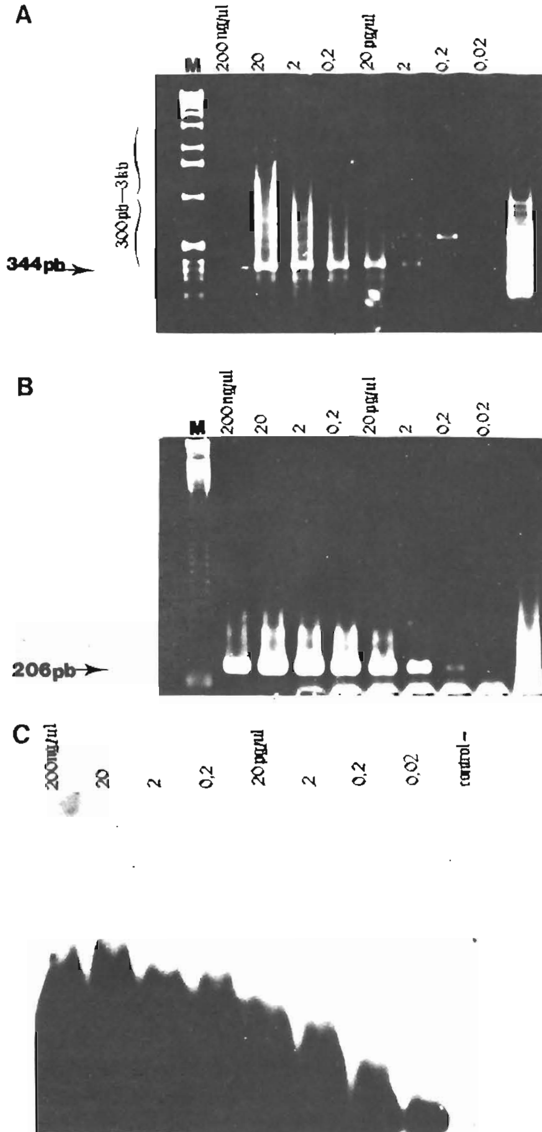


Fig. 2: sensitivity of PCR I and PCR II systems. Electrophoresis in 3% agarose gel of amplified products from serial dilutions of *Plasmodium falciparum* purified DNA. A: PCR I system pf-21; B: PCR III system K114-P1/K114-P2; C: Southern-blot hybridization of B against capture K-14 primer labeled with digoxigenin. The exposure time was 15 min. Lanes - 1: 200 ng/µl; 2: 20 ng/µl; 3: 2ng/µl; 4: 0.2 ng/µl; 5: 20 pg/µl; 6: 2 pg/µl; 7: 0.2 pg/µl; 8: 0.02 pg/µl; M - Fig. 2A: ladder 1 Kb; Fig. 2B: ladder 123 bp.

Effects of anticoagulants on the PCR assays - Previous reports have revealed the inhibitory effect in PCR analysis of certain anticoagulants such as heparin (Barker et al. 1992, Tirasophon et al. 1991), citrate and EDTA (Tirasophon et al. 1994). We tested the PCR assay amplifying vertebrate rDNA included as a positive control for the PCR reaction (Premoli-de-Percoco et al. 1993), taking blood samples with different anticoagulants (heparin and citrate), including a sample in Wathman paper. The results showed the specific vertebrate's rDNA band of 126 bp in all the samples, indicating that there is no inhibition by the anticoagulants (results not shown).

PCR detection of P. falciparum DNA in human blood samples - Fig. 3 shows the agarose 2% electrophoresis of PCR III for a total of 32 samples. There is a 206 bp band (positive signal) in 11 samples (Fig. 3A; lanes 1, 2, 6, 7, 8, 9, 10, 11; Fig. 3B; lanes 2, 10, 14). There was inhibition of the amplification reaction in one sample (Fig. 3B; lane 6) as revealed by the absence of the vertebrate's rDNA amplification band of 126 bp. The same samples were examined using the PCR I system, with the single pf-21 primer (Fig. 4). Those samples

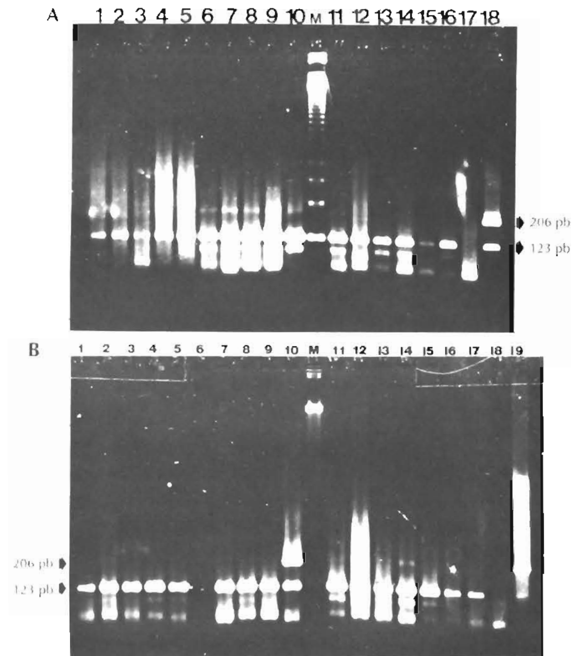


Fig. 3: PCR on human samples, PCR III system. Electrophoresis in 2% agarose gel of amplified products by PCR III. A: lanes 1 to 10 and 11 to 15, human samples 1 to 15; lane 16, negative control with human DNA; lane 17, negative control no DNA; lane 18, positive control with purified *Plasmodium falciparum* DNA. B: lanes 1 to 10 and 11 to 16, human samples 17 to 32; lane 17, negative control with human DNA; lane 18, negative control no DNA; lane 19, positive control with purified *P. falciparum* DNA; M , 123 bp ladder.

with a band pattern similar to the *P. falciparum* DNA positive control (Fig. 4A; lane 18), were considered positives. This positive pattern was observed in 24 samples (Fig. 4A; lanes 1, 2, 5, 6, 7, 8, 9, 10, 12, 13, 14; Fig. 4B; lanes 2, 4, 5, 7, 9, 10, 11, 12, 13, 14, 15, 16, 17). The results showed the specific vertebrate's rDNA band of 126 bp in all the samples, indicating that there is no inhibition of the Taq polimerase by the anticoagulants used.

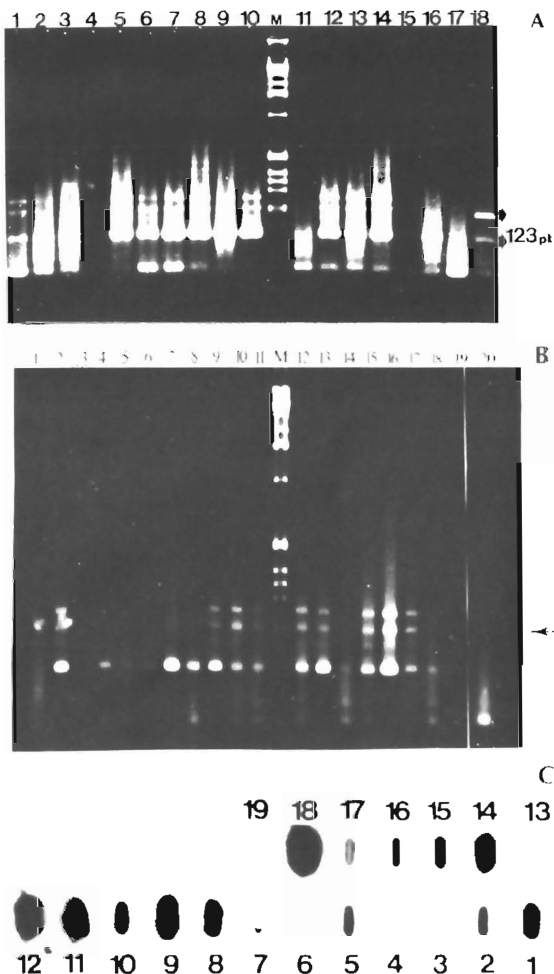


Fig. 4: PCR on human samples, PCR I system. Electrophoresis in 2% agarose gel of amplified products by PCR I pf-21. A: lanes 1 to 15, human samples 1 to 15; lane 16, negative control with human DNA; lane 17, negative control no DNA; lane 18, positive control with purified *Plasmodium falciparum* DNA. B: lanes 1 to 17 are human samples 16 to 32; lane 18, positive control with purified *P. falciparum* DNA; lane 19, negative control with human DNA; lane 20, negative control no DNA; M, 1 Kb ladder. C: slot blot hybridization of amplified products by PCR I pf-21 on human samples 16 to 32 against the pf-21 primer amplified product, labeled with digoxigenin. Exposition time: 30 min. Slots 1 to 17 are human samples 16 to 32; slot 18, positive control with purified *P. falciparum* DNA; slot 19, negative control with human DNA.

Slot-blot hybridization with PCR I pf-21 probe - Fig. 4C shows the slot-blot results of PCR amplifications with PCR I system on human samples. A positive signal was observed in slots 1, 2, 5, 8, 9, 10, 11, 12, 14, 15, 16, and 17.

DNA amplification fingerprint (DAF) of P. falciparum - Six of the 12 Amazonian Brazilian isolates were amplified by PCR I. Fig. 5 shows a different and specific polymorphic fingerprint for each isolate, within the range of 200 bp and 3 Kb. This method could be used as a molecular marker for genetic diversity studies in population genetic.

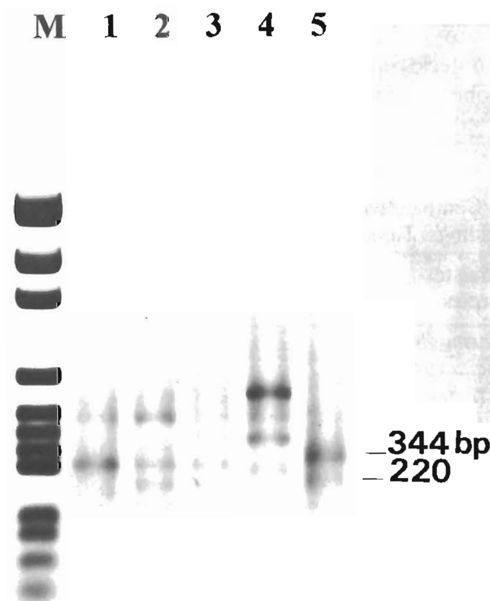


Fig. 5: DAF of *Plasmodium falciparum* isolates from Brazil and Colombia. Electrophoresis in 2% agarose gel of amplified products by PCR I pf-21. Brazilian isolates: 1 - 608; 2 -365; 3 - 51; 4 - 54 and Colombian isolate: 5; M - PGEM molecular weight marker.

Microscopic diagnostic - Table I shows the comparative results of both microscopic readings (100 and 200 fields with immersion lens) from patients blood samples. Detection level was clearly improved when the number of fields was increased. Taking into account this improvement in the microscopic exam sensitivity, we used the 200 fields reading to compare the PCR and hybridization systems. The comparative results of the PCR I, PCR III and microscopy are presented in Table II. Table III illustrates the results of 16 samples derived from the PCR (I) pf-21 test coupled to a hybridization assay, in comparison with the microscopic diagnosis.

DISCUSSION

The current diagnosis of human malaria is achieved by microscopic examination of Giemsa stained blood smears. Although well adapted to the field situation, this methodology is not practical in terms of time and labor involved when large number of samples are required for epidemiological studies. The availability of DNA derived techniques for diagnosis of infectious agents with high specificity and sensitivity has represented an attractive alternative. Several groups have identified species specific repetitive sequences in the genome of *P. falciparum* (Franzen et al. 1984, Oquendo et al. 1986, Barker et al. 1986, McLaughlin et al. 1987) and derived DNA probes in an attempt to improve diagnosis. The highly repeated subtelomeric rep-20 sequence has been target of DNA probes as well as PCR diagnostic assays (Barker et al. 1992). This late approach has increased the detection level but has not significantly surpassed the microscopy diagnosis probably due to self-complementation between the oligonucleotides used.

Exploiting the genomic organization of the pf-21 repeat, we have designed a PCR assay for *P. falciparum* based on a single consensus primer that coupled to non-radioactive hybridization, can be applied to the epidemiological diagnosis of *P. falciparum* with significant improvement over the microscopic analysis.

In order to improve the performance of the diagnostic PCR reaction it was necessary to optimize the protocol for blood sample preparation. The direct treatment of blood with chelex-100 Fe was adopted over the proteinase K incubation and sample preservation on filter paper. Consistent amplification of patient samples was obtained independently of the treatment with anticoagulants, as demonstrated by the vertebrate rDNA PCR assay.

We applied the PCR I system to 32 samples of malaria patients with moderate and low parasitemia. As a reference PCR, we used the PCR III system targeted on moderately repetitive sequences (Fucharoen et al. 1988, Tirasophon et al. 1991) that amplifies an unique band of 206 bp. The sensitivity of these DNA amplification assays was compared to the improved microscopic reading of 200 fields. This more extended examination of the slide increases the positive diagnosis of *P. falciparum* from 24% in 100 fields examined to 38% (Table

TABLE I

Comparative detection levels of two microscopic readings. Tumeremo Bolívar State, Venezuela 1993-94

Parasite species	100 field examined Samples No.	%	200 field examined Samples No.	%
Plasmodium falciparum	7	24	11	38
P. vivax	7	24	6	21
Mixed infection	-	-	2	7
Negatives	15	52	10	34
Total	29	100	29	100

TABLE II

Comparative detection levels between PCR I, PCR III and microscopy. Tumeremo Bolívar State, Venezuela 1993-94

Parasite species	PCR I No. samples	%	PCR III No. samples	%	Microscopy No. samples	%
Plasmodium falciparum	19	66	11	38	11	38
P. vivax	-	-	-	-	6	21
Mixed infection	5	17	5	17	2	7
Negatives	5	17	13	45	10	34
Total	29	100	29	100	29	100

TABLE III

Comparative detection levels between PCR I coupled to an hybridization assay and microscopy. Tumeremo Bolívar State, Venezuela 1993-94

Parasite species	PCR I No. samples	%	PCR III No. samples	%	Microscopy No. samples	%
Plasmodium falciparum	12	75	12	75	7	44
P. vivax	-	-	-	-	2	12
Mixed infection	1	6	2	12	-	-
Negatives	3	19	2	13	7	44
Total	16	100	16	100	16	100

I). Overall, examining the slides at 200 fields favored the diagnosis of mixed infections and the detection of low parasitemia over the 100 field's routine methodology.

Three of the 32 samples showed inhibition of the PCR reaction as evidenced by the absence of the 126 bp product of the vertebrate rDNA PCR assay, and were not include in the comparison analysis. The summary of the results from the 29 patients analyzed by microscopic examination and PCR systems I and III are shown in Table II. The detection ratio of the PCR I system in relation to both the PCR III system and the microscopic examination was 1.7:1. Five of the six diagnosed *P. vivax* infections where detected positive for *P. falciparum* by both PCR I and III systems, increasing the detection of mixed infections more than two fold.

When the PCR I analysis was coupled to hybridization (Fig. 4C) and these results were compared with the microscopic exam (Table III), a similar detection ratio of 1.7:1, with a higher sensitivity for the PCR was observed. The two reported cases of *P. vivax* infection were confirmed positive to *P. falciparum* by hybridization. In general the PCR I system had a higher detection level than both the PCR III system and the 200 field microscopic examination. This increment in the detection of *P. falciparum* infections is significant when sub-patent parasitemia and mixed infections are not detected by the traditional method.

Additionally, the single oligonucleotide amplification produces a DNA fingerprint specific for *P. falciparum* isolates of immediate application to population genetic studies. It was possible to observe a pattern of reproducible bands, specifically a band of 344 bp for the *P. falciparum* isolate from Colombia (Fig. 1A; lane 1, 2; Fig. 5; lane 5). The rep-20 repeat is found in the subtelomeric regions of *P. falciparum* chromosomes (Triglia et al. 1992). It is in this region where chromosome rearrangements responsible for chromosomic size changes observed is postulated to occur. The origin of these size changes is not well understood, but it is attributed to genetic exchange during the different types of cellular division (meiosis-mitosis). The possibility of obtaining an amplification pattern of these regions specific for each isolate (as the result of its chromosomic arrangement), would probably permit to follow in an epidemiological way the different circulating isolates, and the relationship of these DNA amplification fingerprints (DAFs) with other important characters, such as drug resistance and virulence.

In the group of patients studied, it was possible to distinguish at least three DAF variation patterns in the isolates circulating in the zone (Fig. 4A; lanes 1, 2, 8). A totally different pattern was observed in the positive control DNA from Colombia strain (Fig. 4A; lane 18).

These results constitute an important tool in the epidemiology of *P. falciparum* malaria. This approach will allow the analysis of large number of samples for diagnosis, detecting mixed infections and low parasitemia. Concomitantly, circulating isolates can be identified trough specific DAF for *P. falciparum*. We propose the combined use of this polymorphism detection analysis with other similar assays (RAPD, isoenzymes, and micro-satellites) that detect population changes, to further evaluate its discrimination capabilities in identifying different isolates circulating in endemic areas.

ACKNOWLEDGMENTS

To the patients that voluntarily accepted to participate in this study. To Dr Alejandro Caraballo and all the workers in the malaria diagnostic post of the Venezuelan Malaria Program in Tumeremo, Bolívar State, for helping in the sample collection stage. To Dr Marcelo Mazzarri and Dr Alexis Rodriguez for their helpful support. To Dr Moises Wasserman, Dr Hernán Del Portillo and Dr Erlinda Sánchez for providing the reference strain DNAs.

REFERENCES

Barker R, Banchongaksorn T, Courval J, Suwonkerd W, Rimwungtragoon K, Wirth D 1992. A simple method to detect *Plasmodium falciparum* directly from blood samples, using the polymerase chain reaction. *Am J Trop Med Hyg 46*: 416-426.

Barker R, Suebsaeng L, Rooney W, Alecrim G, Dourado H, Wirth D 1986. Specific DNA probe for the diagnosis of *Plasmodium falciparum* malaria. *Science 231*: 1434-1436.

Buesing M, Guerry P, Diesanti C 1987. An oligonucleotide probe for detecting *P. falciparum*: an analysis of clinical specimens from six countries. *J Infect Dis 155*: 1315-1318.

Davis L, Dibner M, Battey J 1986. *Basic Methods in Molecular Biology*. Elsevier, N.Y.

Franzen L, Shabo R, Perlmann H, Wigzell H, Westin G, Aslund L, Persson T, Pettersson U 1984. Analysis of clinical specimens by hybridization with probe containing repetitive DNA from *Plasmodium falciparum*. A novel approach malaria diagnosis. *Lancet I*: 525-527.

Fucharoen S, Tirawanchai N, Wilairat P, Panyim S, Thaithong S 1988. Differentiation of *Plasmodium falciparum* clones by means of a repetitive DNA probe. *Trans R Soc Trop Med Hyg 82*: 209-211.

Lal A, Changkasin S, Holingdadi M, McCutchan T 1989. Ribosomal RNA-based diagnosis of *P. falciparum* malaria. *Mol Biochem Parasitol 36*: 67-72.

Lanar D, McLaughlin G, Wirth D, Barker R, Zolg J, Chulay J 1989. Comparison of thick films, *in vitro* culture and DNA hybridization probes for detecting *P. falciparum* malaria. *Am J Trop Med Hyg 40*: 3-6.

Maniatis T, Fritsch E, Sambrook I 1982. *Molecular Cloning: A Laboratory Manual,* Cold Spring Harbor Laboratory.

McLaughlin G, Breman J, Collins W, Schwartz I, Brandling-Bennet A, Sulzer A, Skinner J, Ruth J, Andrysiak P, Kasaje D, Campbell G 1987. Assessment of a synthetic DNA probe for *Plasmodium falciparum* in African blood specimens. *Am J Trop Med Hyg 37*: 27-36.

Oquendo P, Goman M, Mackay M, Langsley G, Wallıker D, Scaife J 1986. Characterization of repetitive DNA sequence from the malaria parasite, *Plasmodium falciparum*. *Mol Biochem Parasitol 18*: 89-101.

Premoli-de-Percoco G, Pinto-Cisternas J, Ramirez JL, Galindo I 1993. Focal epithelial hiperplasia: Human-papillomavirus induced disease with a genetic predisposition in Venezuela family. *Hum Gent 91*: 386-388.

Tirasophon W, Ponglikitmongkol M, Wilariat P, Boonsaeng N, Panyim S 1991. A novel detection of a single *Plasmodium falciparum* in infected blood. *Biochem Biophys Res Comm 175*: 179-184.

Tirasophon W, Rajkulchai P, Ponglikitmongkol M, Wilairat P, Boonsaeng V, Panyim S 1994. A highly sensitive, rapid and simple polymerase chain reaction-based method to detect human malaria (*Plasmodium falciparum* and *Plasmodium vivax*) in blood samples. *Am J Trop Med Hyg 51*: 308-313.

Triglia T, Wellems T, Kemp D 1992. Towards a high-resolution map of the *Plasmodium falciparum* genome. *Parasitol Today 8*: 225-229.

Wathers AP, McCutchan T 1989. Rapid sensitivity diagnosis of malaria based on ribosomal RNA. *Lancet 8651*: 1343-1346.

WHO - World Health Organization 1997. World malaria situation in 1994. *Weekly Epidemiol Record*, Geneva *72*: 36-269.

Wooden J, Kyes S, Sibley C 1993. PCR and strain identification in *Plasmodium falciparum*. *Parasitol Today 9*: 303-305.

Zolg J, Andrade L, Scott E 1987. Detection of *Plasmodium falciparum* DNA using repetitive DNA clones as species specific probes. *Mol Biochem Parasitol 22*: 145-151.

Systematics and Population Level Analysis of *Anopheles darlingi*

JE Conn

Department of Biology, University of Vermont, Burlington, VT, USA

A new phylogenetic analysis of the Nyssorhynchus subgenus *(Danoff-Burg and Conn, unpub. data) using six data sets {morphological (all life stages); scanning electron micrographs of eggs; nuclear ITS2 sequences; mitochondrial COII, ND2 and ND6 sequences} revealed different topologies when each data set was analyzed separately but no heterogeneity between the data sets using the arn test. Consequently, the most accurate estimate of the phylogeny was obtained when all the data were combined. This new phylogeny supports a monophyletic* Nyssorhynchus *subgenus but both previously recognized sections in the subgenus (Albimanus and Argyritarsis) were demonstrated to be paraphyletic relative to each other and four of the seven clades included species previously placed in both sections. One of these clades includes both* Anopheles darlingi *and* An. albimanus, *suggesting that the ability to vector malaria effectively may have originated once in this subgenus.*

Both a conserved (315 bp) and a variable (425 bp) region of the mitochondrial COI gene from 15 populations of An. darlingi *from Belize, Bolivia, Brazil, French Guiana, Peru and Venezuela were used to examine the evolutionary history of this species and to test several analytical assumptions. Results demonstrated (1) parsimony analysis is equally informative compared to distance analysis using NJ; (2) clades or clusters are more strongly supported when these two regions are combined compared to either region separately; (3) evidence (in the form of remnants of older haplotype lineages) for two colonization events; and (4) significant genetic divergence within the population from Peixoto de Azevedo (State of Mato Grosso, Brazil). The oldest lineage includes populations from Peixoto, Boa Vista (State of Roraima) and Dourado (State of São Paulo).*

Key words: *Anopheles* - *Nyssorhynchus* - *Anopheles darlingi* - phylogeny - combined analysis - parsimony - distance analysis

In the neotropics, species in the subgenus *Nyssorhynchus* are responsible for many of the estimated 20 million annual cases of malaria (Goriup & Pull 1988). The original phylogenetic hypothesis for this important subgenus (Faran 1980, Faran & Linthicum 1981, Linthicum 1988) was based on morphological characters. The subgeneric treatment of Peyton et al. (1992) placed *Nyssorhynchus* species in three purportedly monophyletic sections: Albimanus, Argyritarsis and Myzorhynchella. A recent parsimony analysis of eight species from the Albimanus section using partial sequences of the mtDNA genes ND2 and ND6 (Perera 1993) had no nodes in common with those of Faran (1980). Relationships among species in *Nyssorhynchus* remained unresolved until the recent analysis of Danoff-Burg and Conn (unpub. data) which forms the basis for the systematic portion of this presentation. A subset of their objectives was: (1) an analysis of members of the Albimanus (15 species) and Argyritarsis (8 species) sections using six data sets (morphology, egg ultrastructure, ITS2 region, and mitochondrial genes COII, ND2 and ND6); (2) an examination of character congruence between these data sets using the arn and Templeton tests (Larson 1994, Farris et al. 1995); and (3) a reevaluation of the earlier classifications of *Nyssorhynchus* with regard to a new total evidence phylogeny based on parsimony where all available data were equally weighted and included.

Anopheles darlingi has historically been considered the most important malaria vector throughout much of South America (Deane et al. 1946). Prior to the analysis of Danoff-Burg and Conn (unpub. data), it was placed in the Argyritarsis section of the *Nyssorhynchus* subgenus (Linthicum 1988). It is broadly distributed and has been incriminated as an important regional vector of *Plasmodium falciparum*, the most dangerous of the malaria parasites (Deane et al. 1946), but it is also

These projects received financial support from the National Institutes of Health (NIH; USA) grant AI 31034 to LP Lounibos and NIH grant AI 40116 to JE Conn.
Fax:+802-656.2914. E-mail: jconn@zoo.umv.edu
Received 15 June 1998
Accepted 30 July 1998

a competent vector of other malaria species (Klein et al. 1991). Although the taxonomic status of *An. darlingi* has recently been reevaluated throughout its range and is now considered a single species by allozyme, RAPD-PCR, ITS2 and morphological analysis (Manguin et al. unpub. data), questions remain concerning populations which differ in biting times (reviewed in Rosa-Freitas et al. 1992), life history characteristics (Lounibos et al. 1995) and genetic divergence (discussed in the present work). These heterogeneous traits (and others) may influence the vector competence of *An. darlingi* in different regions in the neotropics. We sequenced two regions (one conserved and one variable) of the mtDNA gene COI from 15 populations of *An. darlingi* from Belize to São Paulo to conduct a population level analysis comparing parsimony and distance methods (Conn & Hennig, unpub. data). Assumptions or hypotheses to be tested at the population level were: (H1) analysis based on distance measurements is more informative than cladistic analysis; (H2) different molecular models (for distance analysis) provide different estimates of genetic divergence; (H3) the more variable region of the mtDNA COI gene gives better resolution among haplotype lineages than the more conserved region; (H4) analysis of combined conserved and variable regions is more informative at various levels of divergence than analysis of either region alone.

MATERIALS AND METHODS

Systematics - Methods of extraction and amplification of DNA, as well as methodology for the phylogenetic analysis of the six data sets, the congruence tests, and outgroup treatment are found in Danoff-Burg and Conn (unpub. data).

Population level - DNA was extracted from individual wild-caught mosquitoes following the protocol in Collins et al. (1987). These mosquitoes were collected from BZ (Belize), EJ (El Juval, Trujillo, Venezuela), AY (Puerto Ayacucho, Amazonas, Venezuela), FG (French Guiana), IQ (Iquitos, Peru), GU (Guayaramerín, Bolivia), BV (Boa Vista, Roraima, Brazil), MC (Macapá, Amapá, Brazil), CP (Capanema, Pará, Brazil), NS (Tefé, Amazonas, Brazil), IT (Itacoatiara, Amazonas, Brazil), AB (Porto Velho, Rondônia, Brazil), PX (Peixoto de Azevedo, Mato Grosso, Brazil), SP (Araraquara, São Paulo, Brazil) and DO (Dourado, São Paulo, Brazil) (Fig.). Primers for the two regions of the COI gene were from Lunt et al. (1996). For the parsimony analysis the following parameters were used with PAUP 3.1.1. (Swofford 1993): unrooted trees, heuristic search, and trees were assessed by bootstrap analysis (100-1000 replicates) (Felsenstein 1985). Three distance models were assessed by the Neighbor-Joining (NJ) distance method: uncorrected p, Tamura-Nei and HKY85 and all were tested by 1000 bootstrap replications. Details of this analysis are from Conn and Hennig (unpub. data).

Collection localities for samples of *Anopheles darlingi*

RESULTS

Systematics - The reanalysis of the 49 morphological characters from Faran (1980), Faran and Linthicum (1981) and Linthicum (1988) produced a largely unresolved strict consensus tree which was eight steps shorter than the original (Danoff-Burg & Conn, unpub. data). The analysis of the six different data sets (morphology, egg ultrastructure, ITS2 region, and mitochondrial genes COII, ND2 and ND6) each produced a distinctive topology even though the character-based heterogeneity was not significantly different between any two matrices. The tree based on the ITS2 analysis alone most accurately reflected the new phylogenetic hypothesis which was produced when all data were combined. This new phylogeny supported monophyly of the subgenus *Nyssorhynchus* but both the Albimanus and Argyritarsis sections were paraphyletic relative to each other and four clades contained species previously placed in both sections. Both *An. albimanus* (previously in the Albimanus section) and *An. darlingi* (previously in the Argyritarsis section) arose basally in the phylogeny.

Population level - For the 27 individuals sequenced for the conserved region alone (315 bp), there were 20 characters that were parsimony informative. With 1,000 bootstrap replicates, four clades were supported: BV-NS at 53%; CP1-CP2-DO at 66%; FR-MC at 53%; PX1-PX2 at 100%.

Individuals from the same locality were found in the same clade 50% of the time and little geographic partitioning was apparent. For the variable region alone (425 bp), 66 characters were parsimony informative among the 24 individuals sequenced.

Two clades were supported (100 bootstrap replicates): EJ1-GU1-EJ2-DO1-NS4-AB3 at 52% and SP2-PX1-BV3-PX2-DO2-NS3-AB4 at 61%. Except for PX and EJ, individuals from the same populations were not found in the same clades. The combined region (740 bp) had 81 characters that were parsimony informative for the 22 individuals sequenced. Three clades were supported at 200 bootstrap replications: EJ1-GU1-EJ3 at 68%; PX1-PX2-BV3-DO2 at 100% and NS4-AB3 at 82%. The clade composition was similar to that found in the variable region even though one additional clade was supported by the analysis of the combined regions. The parsimony trees for each of the COI regions were tested to determine if there was phylogenetic signal by comparing each of the three tree lengths {conserved region (TL = 53), variable region (TL = 256), combined regions (TL =289)} to a distribution of tree lengths of 1000 randomly generated trees. All three trees contained significant phylogenetic signal using this test.

A comparison of the parsimony analysis of the combined regions with each of the three distance models demonstrated that the trees from the parsimony and the HKY85 model were identical in clade composition (EJ1-GU1-EJ3, PX1-PX2-BV3-DO2, NS4-AB3) and differed very slightly in levels of bootstrap support. For the uncorrected p and Tamura-Nei models, three additional populations were supported (SP2, AB4, and NS3) either as a separate clade (uncorrected p) or as part of a larger clade (Tamura-Nei).

Results of the four hypotheses were: (H1) trees were either more resolved using parsimony (conserved and variable regions analyzed separately) or nearly equally resolved (combined analysis) as compared with distance analyses; (H2) of the three distance models tested, HKY85, the most parameter-rich model, gave the least amount of resolution (i.e., lowest levels of bootstrap support); (H3) for conserved vs. variable regions of the COI gene, parsimony analysis resulted in more lineages being supported (four) for the conserved region compared with the variable region (two) but most of the same populations were grouped with both regions; while for the distance analyses between regions, each model contained the same number of lineages; (H4) the combined analysis for both the parsimony and distance was more informative than for either region alone.

A graph of the frequencies of pairwise genetic distances of both regions combined using uncorrected p (x-axis) compared with the number of pairwise comparisons (y-axis) resulted in a bimodal distribution. The genetic distances for the first peak ranged from 0.004-0.05. All of the pairwise comparisons of high genetic distances (0.06-0.13; the second peak) included at least one individual from PX, BV or DO.

DISCUSSION

Systematics - The parsimony analysis of both molecular and morphological data does not support the earlier phylogenies of Faran (1980), Faran and Lithicum (1981), Linthicum (1988) and Perera (1993). The discordant topologies of the six data sets were probably the result of two factors: (1) homoplasy, and (2) data sets were informative at a different taxonomic levels. Perhaps the most significant aspect of this new phylogenetic hypothesis is the basal position of two of the major neotropical malaria vectors, *An. albimanus* and *An. darlingi*, suggesting that the ability to effectively vector malaria parasites may have arisen once in the ancestor of *Nyssorhynchus* (Danoff-Burg & Conn, unpub. data). This pleisiotypic ability appears to have been retained by many species in this subgenus which act as important local or regional vectors when population densities of major vectors are low or when environmental conditions are significantly altered (Cruz Marques 1986, Póvoa et al. unpub. data). The similarity between the ITS2 tree and the total evidence tree suggests that nuclear markers are potentially more accurate in reconstructing the true phylogeny at this hierarchical level compared with the mitochondrial or morphological markers presented.

Population level - The similar results between the parsimony and distance models were presumably because there was low overall sequence divergence among populations (i.e., not near saturation), there were no secondary hits, and many of the mutations were unique to single individuals.

If the genetic distances and lineage support are accurate portrayals of the evolutionary history of *An. darlingi*, this suggests that there have been at least two waves of colonization events across South America. Alternatively, the haplotype lineage with the greatest genetic divergence, PX-BV-DO, may be the remnant of an older lineage which has gone extinct in other regions of the range of *An. darlingi*; this scenario is favoured by the strong support for the PX-BV-DO lineage and how rarely individuals from the same geographic locality are found in the same clade. This pattern of distribution (multiple divergent haplotypes in the same population) has also been found in *An. nuneztovari* (Conn et al. 1998) and may be more appropriately explained by local extinction of once widespread lineages

than by heterogeneous effective population sizes or immigration from previously isolated areas (Slatkin 1985).

Information on genetic divergence and heterogeneity among populations of *An. darlingi* may be useful for existing malaria control strategies in that local solutions will need to be implemented for successful transmission reduction, as proposed by WHO (1992). One of the worst recent malaria outbreaks in Brazil has been documented in Peixoto de Azevedo (R. Zimmermann, pers. comm.) where *An. darlingi* is considered to be the main vector.

ACKNOWLEDGMENTS

For logistic and technological support in the field I thank collaborators in Bolivia (Unidad Sanitaria, Riberalta), Brazil (Fundação Nacional da Saúde, Fundação Oswaldo Cruz, Institituo Evandro Chagas, Instituto de Pesquisas Científicas e Tecnológicas do Estado do Amapá), France (LIN-ORSTOM), Peru (NAMRID, Iquitos) and Venezuela (Univ. de los Andes, Trujillo; Univ Central, Caracas; Malariología, Maracay). To RC Wilkerson for additional specimens from Brazil. To WC Kilpatrick for constructive criticism. To JA Danoff-Burg and A Hennig for the excellent collaboration.

REFERENCES

Collins FH, Mendez MA, Rasmussen MO, Mehaffey PC, Besansky NJ, Finnerty V 1987. A ribosomal RNA gene probe differentiates member species of the *Anopheles gambiae* complex. *Am J Trop Med Hyg 37*: 37-41.

Conn JE, Mitchell SE, Cockburn AF 1998. Mitochondrial DNA analysis of the neotropical malaria vector *Anopheles nuneztovari*. *Genome*, in press.

Cruz Marques A 1986. Um estudo sobre a dispersão de casos de malária no Brasil. *Rev Bras Malar D Trop 38*: 51-75.

Deane LM, Causey OR, Deane MP 1946. Studies on Brazilian anophelines from the Northeast and Amazon regions. I. An illustrated key by adult female characteristics for the identification of thirty-five species of Anophelini, with notes on the malaria vectors (Diptera, Culicidae). *Am J Hyg Mono Series 18*: 1-18.

Faran ME 1980. Mosquito studies (Diptera, Culicidae) XXXIV. A revision of the Albimanus section of the subgenus *Nyssorhynchus* of *Anopheles*. *Contr Amer Ent Inst 15*: 1-215.

Faran ME, Linthicum KJ 1981. A handbook of Amazonian species of *Anopheles* (*Nyssorhynchus*) (Diptera: Culicidae). *Mosq Syst 8*: 1-107.

Farris JS, Kallersjo M, Kluge AG, Bult C 1995. Constructing a significance test for incongruence. *Syst Biol 44*: 570-572.

Felsenstein J 1985. Confidence limits on phylogenies: an approach using the bootstrap. *Evolution 39*: 783-791.

Goriup S, Pull JH 1988. Field research in the context of malaria control, p. 1741-1764. In WH Wernsdorfer, I McGregor (eds), *Malaria: Principles and Practice of Malariology*, Vol. 2, Churchill and Livingstone, New York.

Klein TA, Lima JBP, Tada MS, Miller R 1991. Comparative susceptibility of anophelines in Rondônia, Brazil to infection by *Plasmodium vivax*. *Am J Trop Med Hyg 45*: 463-470.

Larson A 1994. The comparison of morphological and molecular data in phylogenetic systematics, p. 371-390. In G Schierwater, B Streit, GP Wagner, R DeSalle (eds), *Molecular Ecology and Evolution: Approaches and Applications*, Birkhauser, Verlag Basel/Switzerland.

Linthicum KJ 1988. A revision of the Argyritarsis section of the subgenus *Nyssorhynchus* of *Anopheles*. *Mosq Syst 20*: 99-271.

Lounibos LP, Nishimura N, Conn J, Lourenço-de-Oliveira R 1995. Life history correlates of adult size in the malaria vector *Anopheles darlingi*. *Mem Inst Oswaldo Cruz 90*: 769-774.

Lunt DH, Zhang DX, Szymura JM, Hewitt GM 1996. The insect cytochrome oxidase I gene: evolutionary patterns and conserved primers for phylogenetic studies. *Insect Mol Biol 5*: 153-165.

Perera OP 1993. *Phylogenetic Analysis of Two Mitochondrial Genes from Several Species of the Subgenus Nyssorhynchus (Culicidae: Anopheles) and the Development of Species-specific DNA Probes for Their Identification*, PhD Thesis, University of Florida, Gainesville, 110 pp.

Peyton EL, Wilkerson RC, Harbach RE 1992. Comparative analysis of the subgenera *Kerteszia* and *Nyssorhynchus* of *Anopheles* (Diptera: Culicidae). *Mosq Syst 24*: 51-69.

Rosa-Freitas MG, Broomfield G, Priestman A, Milligan JJM, Momen H, Molyneux DH 1992. Cuticular hydrocarbons, isoenzymes, and behavior of three populations of *Anopheles darlingi* from Brazil. *J Amer Mosq Control Assoc 8*: 357-366.

Slatkin M 1985. Gene flow in natural populations. *Ann Rev Ecol Syst 16*: 393-430.

Swofford DL 1993. PAUP: phylogenetic analysis using parsimony. Version 3.1.1. Illinois Natural History Survey, Champaign.

WHO - World Health Organization 1992. World malaria situation in 1990. *Bull WHO 70*: 801-807.

Anopheline Species Complexes in Brazil. Current Knowledge of Those Related to Malaria Transmission

Maria Goreti Rosa-Freitas/+, Ricardo Lourenço-de-Oliveira*, Carlos José de Carvalho-Pinto*/, Carmen Flores-Mendoza*, Teresa Fernandes Silva-do-Nascimento***

Laboratório de Sistemática Bioquímica, Departamento de Bioquímica e Biologia Molecular *Laboratório de Transmissão de Hematozoários, Departamento de Entomologia, Instituto Oswaldo Cruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil **Departamento de Microbiologia e Parasitologia, Universidade Federal de Santa Catarina, Florianópolis, SC, Brasil

A summary of the problems related to the systematics of primary and secondary Brazilian anophelines vectors of malaria is presented.

Key words: *Anopheles* systematics - species complexes - malaria vectors - *Nyssorhynchus* - *Kerteszia*

Many neotropical anopheline species are either candidates or formed by complex of cryptic species. The taxonomic elucidation of these complexes reflects on the epidemiology of malaria transmission and ultimately to the control.

In Brazil, there are 54 species belonging to five subgenera of *Anopheles* Meigen (*Nyssorhynchus, Kerteszia, Stethomyia, Lophopodomyia, Anopheles*). Anopheline species reported as human malaria vectors in the country belong to the subgenera *Nyssorhynchus* and *Kerteszia* (Deane 1986, Consoli & Lourenço-de-Oliveira 1994).

In the subgenus *Nyssorhynchus*, the species found harboring human plasmodia include *Anopheles darlingi* Root 1926, *An. aquasalis* Curry 1932, *An. albitarsis sensu lato* Lynch-Arribálzaga 1878 (including *An. deaneorum* Rosa-Freitas 1989), *An. oswaldoi* Peryassú 1922, *An. nuneztovari* Gabaldon 1940 and *An. triannulatus* (Neiva & Pinto 1922). In the subgenus *Kerteszia* natural infections were reported for *An. cruzii* Dyar & Knab 1908, *An. bellator* Dyar & Knab 1906 and *An. homunculus* Komp 1937. It is our opinion that other species reported naturally infected do not play a role in malaria maintenance as they are exophilic, zoophilic, of low density and their distribution and frequency do not coincide with that of malaria. Except for *An. darlingi*, the natural history of the species listed above points out for zoophilic and/or exophilic behavior in some areas, in such a fashion that their role in malaria transmission is doubted (Deane 1986). Are these characteristics an indication that these species are indeed complexes?

To decide whether a given species is high polymorphic or a complex of closely related species, integrated approach studies on distinct populations, including on that of the type-localities and where morphological/behavioral/molecular differences have been reported, are mandatory. Most of the Brazilian anopheline species has been taxonomically investigated by morphology, behavior and molecular tools such as isoenzymes and DNA analyses (mitochondrial and ribosomal DNA restriction analysis, random amplification and sequencing of specific regions) as summarized on Table.

More than 99% of the malaria cases reported in Brazil occur in the Amazon in which transmission is due to *Nyssorhynchus* species only.

An. darlingi is the most important Brazilian malaria vector (Shannon 1933, Rachou 1958). The species is the most anthropophilic and endophilic among the Amazonian anophelines. It is frequently found infected and its distribution and density are clearly related to malaria transmission. Even though many populations of the species have been lately reported as biting outdoors, *An. darlingi* continues successfully transmitting malaria both indoors and at the close vicinity of the houses (Lourenço-de-Oliveira 1995). Isoenzymatic, behavioral and mitochondrial DNA studies on either Brazilian (Rosa-Freitas et al. 1992, Freitas-Sibajev et al. 1995) or other Latin-American (Manguin et al. 1998) populations, showed that *An. darlingi* is a monotypic species.

An. aquasalis is the lowland coastal vector in Brazil. Chromosomal banding pattern and mtDNA

+Corresponding author.
E-mail: mgoreti@gene.dbbm.fiocruz.br
Received 15 June 1998
Accepted 30 July 1998

TABLE

Summary of differences in behavior, morphology, isoenzyme, mtDNA, rDNA, RAPD and cytogenetic data reported in the literature for populations of neotropical anopheline species related to malaria transmission

Species	Behavior	Morphology	Isoenzyme	mtDNA	rDNA (ITS2)	RAPD	Cytogenetics	References	Conclusion
Anopheles darlingi	≠ (Peak and place of biting)	≅ (Except from Belize)	≅	≅	ND	ND	≠	Consoli & Lourenço-de-Oliveira 1994 Freitas-Sibajev et al. 1995 Harbach et al. 1993 Kreutzer et al. 1972 Manguin et al. 1998 Rosa-Freitas et al. 1992	Monotypic
An. aquasalis	≠ (Host and place of biting)	≠ (Egg)	≅	≠	≅	ND	≅	Conn et al. 1993a Cova-Garcia et al. 1977 Flores-Mendonza 1994 Monaca-Perez & Conn 1991	High polymorphic (UI)
An. albitarsis	≠ (Host and place of biting)	≅ (Except *An. deaneorum*)	≠	≠	ND	≠	≠	Kreutzer et al. 1976 Narang et al. 1993 Rosa-Freitas et al. 1990 Wilkerson et al. 1995	Complex (4 species)
An. oswaldoi	≠ (Host and place of biting)	≠ (Male genitalia)	UI	ND	≠	ND	ND	Causey et al. 1946 Consoli & Lourenço-de-Oliveira 1994 Flores-Mendoza pers. comm. Klein & Lima 1990 Marrelli et al. 1998	Complex: (at least 2 forms - UI)
An. nuneztovari	≠ (Host and place of biting)	≠ (Egg, male genitalia and female)	≠	≠	≠	ND	≠	Delgado & Rubio-Palis 1992 Fritz et al. 1994 Hribar 1994, 1995 Linley et al. 1996	Possibly a complex
An .triannulatus	≅	≠ (Male genitalia, larva and egg)	≠	≠	ND	≠	ND	Silva-do-Nascimento1995 Silva-do-Nascimento pers. com.	Complex: (at least 2 species - UI)
An. cruzii	≠ (Acrodendrophily)	≠ (Larva)	ND	ND	≠	ND	≠	Deane et al. 1971 Malafronte et al. 1997 Ramirez et al. 1989 Zavortink 1973	High polymorphic (UI)

ND: non determined; UI: under investigation.

restriction profiles of specimens from Venezuela and Brazil were identical (Moncada-Pérez & Conn 1991, Conn et al. 1993a). Isoenzymes from three populations of Venezuela and Surinam (Steiner et al. 1981) and two from Brazil (Flores-Mendoza 1994) with behavioral differences also revealed only intraspecific variation. Egg morphology of *An. aquasalis* varies intraspecifically (Maldonado et al. 1997). In fact, variation was seen in a single female oviposition (Flores-Mendoza 1994). Results of mitochondrial DNA and egg morphology analyses however, suggest that there might be an interspecific division in *An. aquasalis* populations north and south of the Amazon River delta (Conn et al. 1993a, Linley et al. 1993).

An. albitarsis is a complex formed by, at least, four sibling species: *An. marajoara* Galvão & Damasceno 1942, *An.albitarsis sensu stricto* Lynch-Arribálzaga 1878, *An. deaneorum* and a fourth form to be formally described (Rosa-Freitas et al. 1990, Wilkerson et al. 1995). Since *An. deaneorum* is the only morphologically distinguishable member of the albitarsis complex, the role of each member in malaria transmission has not been determined yet. The incrimination of other members of the complex, except *An. deaneorum* (Klein et al. 1991a, b), were based solely on their presumed geographical distribution.

An. oswaldoi has been regarded as a potential malaria vector in some localities of the Amazon (Arruda et al. 1986, Oliveira-Ferreira et al. 1990, Branquinho et al. 1996) although some authors believe that most populations of this species are much more related to the natural environment and prefer to feed on animals than on man indoors (Deane et al. 1948, Consoli & Lourenço-de-Oliveira 1994, Lourenço-de-Oliveira & Luz 1996). The taxonomic status of the species is under investigation (PhD work of CFM). Preliminary results from morphological analyses demonstrate that at least two forms are present under An. *oswaldoi*: *An. oswaldoi sensu stricto* and *An. konderi* Galvão & Damasceno 1942, distinguished pratically only by the shape of the apex of aedeagus (Causey et al. 1946, Lounibos et al. 1997).

Morphology, behavior, cytogenetics, isoenzymes and mtDNA studies favor the existence of at least two cryptic species in *An. nuneztovari*: one in Venezuela and Colombia northwest of Orinoco and another in the Amazon (Conn et al. 1993b, Fritz et al. 1994, Linley et al. 1996). The species is considered a primary malaria vector in Venezuela and Colombia (Gabaldon 1969, Gabaldon et al. 1975). In Brazil however, the species is not related to malaria transmission, although natural infection by *Plasmodium vivax* has been detected in areas where *darlingi* was the primary vector (Arruda et al. 1986).

An. triannulatus is constituted by at least three sibling forms. These forms can be differentiate morphologically (egg, larva and male genitalia) and isoenzymatically (Silva-do-Nascimento 1995). The typical *triannulatus* is the most known and largely distributed form. The other two forms seem to be restricted mostly to central Brazil and are not related to malaria transmission.

The mosquitoes of the subgenus *Kerteszia* share the common characteristic of using bromeliads as breeding places. An exception is *An.* (*Ker.*) *bambusicolus* Komp 1937 that also breeds in bamboo.

An. (*Kersteszia*) *cruzii* and *An. bellator* were primary vectors of the malaria once endemic in southeastern/southern Brazil (Rachou 1958). *An. cruzii* is currently involved in the maintenance of the oligosymptomatic malaria occurring in the valleys of the Atlantic Coastal Rain Forest in both Rio de Janeiro and São Paulo states (Carvalho et al. 1988, Azevedo 1997, Branquinho et al. 1997). Larval differences were observed in *An. cruzii* populations from Rio de Janeiro and Santa Catarina (Zavortink 1973). Besides, chromosomal banding pattern differences were also found among several *An. cruzii* populations (Ramirez 1989, Dessen pers. comm.). *An. homunculus* is a morphologically close related species and there is the possibility of being a sibling species in the cruzii complex (PhD work of CJCP). The remaining *Kerteszia* species do not seem to be important in malaria transmission in Brazil.

In summary, *An. darlingi* is a monotypic species. *An. aquasalis* and *An. nuneztovari* are possibly complexes. *An. albitarsis*, *An. triannulatus* and *An. oswaldoi* are complexes of species. Anopheline species of the subgenus *Kersteszia* are still under investigation (Table).

The refinement of the taxonomic tools and the addition of other populations are likely to lead to new insights into the knowledge and understanding of the neotropical species complexes.

REFERENCES

Arruda ME, Carvalho MB, Nussenzweig RS, Maracic M, Ferreira AW, Cochrane AH 1986. Potential vectors of malaria and their different susceptibility to *Plasmodium falciparum* and *P. vivax* in northern Brazil identified by immunoassay. *Am J Trop Med Hyg 35:* 873-881.

Azevedo AL 1997. *Aspectos da Epidemiologia da Malária e da Biologia de Anopheles (Kerteszia) cruzii Dyar & Knab em Vales Montanhosos do Sistema de Mata Atlântica*, MSc Thesis, Instituto Oswaldo Cruz , Fiocruz, Rio de Janeiro, 94 pp.

Branquinho MS, Araujo MS, Natal D, Marrelli MT, Rocha RM, Taveira FA, Kloetzel JK 1996. *Anoph-*

eles oswaldoi a potential malaria vector in Acre, Brazil. *Trans R Soc Trop Med Hyg 90*: 233.

Branquinho MS, Marrelli MT, Curado I, Natal D, Barata JM, Tubaki R, Carreri-Bruno GC, de Menezes RT, Kloetzel JK 1997. Infecção de *Anopheles* (*Kerteszia*) *cruzii* por *Plasmodium vivax* e *Plasmodium vivax* variante VK247 nos municípios de São Vicente e Juquitibá, São Paulo. *Rev Panam Salud Pública 2*: 189-193.

Carvalho ME, Glasser CM, Ciaravolo RMC, Etzel A, Santos LA, Ferreira CS 1988. Sorologia de malaria vivax no foco Aldeia dos Índios, município de Peruíbe, São Paulo, de 1984 a 1986. *Cad Saúde Públ 4*: 276-292.

Causey OR, Deane LM, Deane MP 1946. Studies on Brazilian anophelines from the northeast and Amazon regions. II. An illustrated key by male genitalia characteristics for the identification of thirty-four species of anophelini, with a note on dissection technique. *Am J Hyg Mono Series 18*: 21-31.

Conn JE, Cockburn AF, Mitchell SE 1993a. Population differentiation of the malaria vector *Anopheles aquasalis* using mitochondrial DNA. *J Hered 84*: 248-253.

Conn JE, Rangel-Puertas Y, Seawright J 1993b. A new cytotype of *Anopheles nuneztovari* from Western Venezuela and Colombia. *J Am Mosq Control Assoc 9*: 294-301.

Consoli R, Lourenço-de-Oliveira R 1994. *Principais Mosquitos de Importância Sanitária no Brasil*, Ed. Fiocruz. 225 pp.

Cova-Garcia P, Pulido J, Amarista RJ 1977. *Anopheles (Nyssorhynchus) deltaorinoquensis* n. sp. (Diptera: Culicidae) de Venezuela. *Bol Dir malariol San Amb 17*: 150-160.

Deane LM 1986. Malaria vectors in Brazil. *Mem Inst Owaldo Cruz 81*(Suppl.II): 5-14.

Deane LM, Causey OR, Deane MP 1948. Notas sobre a biologia do anofelinos das regiões nordestina e amazônica do Brasil. *Rev Serv Esp Saúde Publ 1*: 827-965.

Deane LM, Deane MP, Ferreira-Neto JA, Almeida FB 1971. On the transmission of simian malaria in Brazil. *Rev Inst Med Trop São Paulo 13*: 311-319.

Delgado N, Rubio-Paris Y 1992. Morphometric characterization of the malaria vector *Anopheles nuneztovari* (Diptera: Culicidae) from Western Venezuela. *Mosq Syst 24*: 241.

Flores-Mendoza C 1994. *Um Estudo sobre Alguns Aspectos Comportamentais, Morfológicos e Bioquímicos de Anopheles aquasalis Curry 1932*, MSc Thesis, Instituto Oswaldo Cruz, Rio de Janeiro, 172 pp.

Freitas-Sibajev MGR, Conn J, Mitchell S, Cockburn AF, Seawright JA, Momen H 1995. Mitochondrial DNA and morphological analyses of *Anopheles darlingi* populations from Brazil. *Mosq Syst 27*: 79-85.

Fritz GN, Conn J, Cockburn A, Seawright J 1994. Sequence analysis of the ribosomal DNA internal transcribed spacer 2 from populations of *Anopheles nuneztovari*. *Mol Biol Evol 11*: 406-416.

Gabaldon A 1969. Global malaria erradication: changes of strategy and future outlook. *Am J Trop Med Hyg 18*: 641-656.

Gabaldon A, Martin GG, Sifontes R 1975. Necesidades en el campo de la investigacion del Programa Nacional de Eradicacion y Control de la Malaria de Venezuela. *Bol Dir Malariol San Amb 15*: 263-285.

Harbach RE, Roberts DR, Manguin S 1993. Variation in the hindtarsal markings *of Anopheles darlingi* (Diptera: Culicidae) in Belize. *Mosq System 25*: 192-197.

Hribar LJ 1994. Geographic variation of male genitalia of *Anopheles nuneztovari* (Diptera: Culicidae). *Mosq Syst 26*: 132-144.

Hribar LJ 1995. Costal wing spot variation within and among progeny of single female *Anopheles nuneztovari* (Diptera: Culicidae). *Mosq Syst 27*: 1-15.

Klein TA, Lima JBP 1990. Seazonal distribution and biting patterns of *Anopheles* mosquitoes in Costa Marques, Rondônia, Brazil. *J Am Mosq Control Assoc 6*: 700-707.

Klein TA, Lima JBP, Tada MS 1991a. Comparative susceptibility of anopheline mosquitoes to *Plasmodium falciparum* in Rondônia, Brazil. *Am J Trop Med Hyg 44:* 598-603.

Klein TA, Lima JBP, Tada MS, Miller R 1991b. Comparative susceptibility of anopheline mosquitoes in Rondônia, Brazil, to infection by *P. vivax*. *Am J Trop Med Hyg 45:* 463-470.

Kreutzer RD, Kitzmiller JB, Ferreira E 1972. Inversion polymorphism in the salivary gland chromosomes of *Anopheles darlingi*. *Mosq News 32*: 555-556.

Kreutzer RD, Kitzmiller JB, Rabbani MG 1976. Cytogenetically distinguishable populations of the mosquito *Anopheles albitarsis*. *Acta Amazonica 6*: 473-481.

Linley JR, Lounibos LP, Conn J 1993. A description and morphometric analysis of the eggs of four South American populations of *Anopheles aquasalis*. *Mosq Syst 25*: 198-214.

Linley JR, Lounibos LP, Conn J, Duzak D, Nishimura N 1996. A description and morphometric comparison of eggs from eigth geographic populations of the South American malaria vector *Anopheles* (*Nyssorhynchus*) *nuneztovari*. *J Am Mosq Control Assoc 12*: 275-292.

Lounibos LP, Duzak D, Linley JR 1997. Comparative egg morphology of six species of the Albimanus Section of *Anopheles* (*Nyssorhynchus*) (Diptera: Culicidae). *J Med Entomol 34*:136-155.

Lourenço-de-Oliveira R 1995. Qual a importância da hematofagia extradomiciliar do *Anopheles darlingi* na Amazônia? *Rev Patol Trop 23* (suppl.): 100-101.

Lourenço-de-Oliveira R., Luz SLB 1996. Simian malaria at two sites in the Brazilian Amazon. II. Vertical distribution and frequency of anopheline species inside and outside the forest. *Mem Inst Oswaldo Cruz 91*: 687-694.

Malafronte RS, Marrelli MT, Carreri-Bruno, GC, Urbinatti, PR, Marinotti O 1997. Polymorphism in the second internal transcribed spacer (ITS2) of *Anopheles (Kerteszia) cruzi* (Diptera: Culicidae)

from the State of São Paulo, Brazil. *Mem Inst Oswaldo Cruz 92* (Suppl. I): 306.

Maldonado V, Finol HJ, Navarro JC 1997. *Anopheles aquasalis* eggs from two Venezuelan localities compared by scanning electron microscopy. *Mem Inst Oswaldo Cruz 92*: 487-491.

Manguin S, Wilkerson RC, Conn JE, Rubio-Palis Y, Danoff-Burg JA, Roberts DR 1998. Population structure of the primary malaria vector in South America, *Anopheles darlingi* using isozyme, RAPD, ITS2 and morphological markers. *Am J Trop Med Hyg*, in press.

Marrelli MT, Malafronte RS, Flores-Mendoza C, Lourenço-de-Oliveira R, Kloetzel JK, Marinotti O 1998. Polymorphism in the second internal transcribed spacer (ITS2) of ribosomal DNA among specimens of *Anopheles oswaldoi*. *Insect Mol Biol*, in press.

Moncada-Pérez A, Conn J 1991. A polytene chromosome study of four populations of *Anopheles aquasalis* from Venezuela. *Genome 35*: 327-331.

Narang SK, Klein TA, Perera OP, Lima JB, Tang AT 1993. Genetic evidence for the existence of cryptic species in the *Anopheles albitarsis* complex in Brazil: allozymes and mitochondrial DNA restriction fragment length polymorphisms. *Biochem Genet 31*: 97-112.

Oliveira-Ferreira J, Lourenço-de-Oliveira R, Teva A, Deane LM, Daniel-Ribeiro CT 1990. Natural malaria infections in anophelines in Rondonia State, Brazilian Amazon. *Am J Trop Med Hyg 43*: 6-10.

Rachou RG 1958. Anofelinos do Brasil. Comportamento das espécies vetoras de malária. *Rev Bras Malariol Doenças Trop 10:* 145-181.

Ramirez CL 1989. *Estudo Cromossômico em uma População de Anopheles (Kerteszia) cruzii Dyar & Knab 1909,* MSc Thesis, Instituto de Biociências, Universidade de São Paulo, São Paulo, 96 pp.

Rosa-Freitas MG, Broomfield G, Priestmann A, Milligan P, Momen H, Molyneux DH 1992. Studies on cuticular components, isoenzymes and behaviour of 3 populations of *Anopheles darlingi* from Brazil. *J Amer Mosq Control Assoc 8*: 357-366.

Rosa-Freitas MG, Deane L, Momen H 1990. A morphological, behavioural and isoenzymatic study in *Anopheles albitarsis* from 10 populations. *Mem Inst Oswaldo Cruz 85*: 275-289.

Shannon RC 1933. Anophelines of the Amazon Valley. *Proc Entomol Soc Washington 35*: 117-143.

Silva-do-Nascimento T 1995. *Estudo Taxonômico e Notas sobre a Biologia de Anopheles triannulatus de 8 Localidades,* MSc Thesis, Instituto Oswaldo Cruz, Fiocruz, Rio de Janeiro, 88 pp.

Steiner WWM, Kitzmiller JB, Osterbur, DL 1981. On the genetic identity and evolution of the malaria vectors *Anopheles aquasalis* and *Anopheles emilianus*. In R Pal, JB Kitzmiller & T Kanada (eds), *Cytogenetics and Genetics of Vectors*. Proceedings of a symposium of the XVI International Congress of Entomology, Amsterdam, Elsevier Biomedical, Amsterdam.

Wilkerson RC, Parsons TJ, Klein TA, Gaffigan TV, Bergo E, Consolim J 1995. Diagnosis by random amplified polymorphic DNA polymerase chain reaction of four cryptic species related to *Anopheles albitarsis* from Paraguay, Argentina and Brazil. *J Med Entomol 32*: 697-704.

Zavortink TJ 1973. Mosquito studies XXIX. A review of the subgenus *Kerteszia* of *Anopheles*. *Contr Amer Entomol Inst 9*: 1-54.

Implications of a Neotropical Origin of the Genus *Leishmania*

Harry Noyes

Liverpool School of Tropical Medicine, Pembroke Place, Liverpool, L3 5QA, UK

The hypothesis of a Neotropical origin of the Leishmania/Endotrypanum *clade is reviewed. The position of the* L. (Sauroleishmania) *external to the subgenus* L. (Leishmania) *is not consistent with the Neotropical origin of the latter subgenus. It is suggested that this may be a consequence of a faster evolutionary rate in the* L. (Sauroleishmania). *The implications for the classsification of the phlebotomine sandflies of the hypothesis for a Neotropical origin of the* Leishmania *is also considered. The classification of Galati (1995) is proposed to be most consistent with the hypothesis of a Neotropical origin of the* Leishmania, *whilst classifications which place the New and Old World species in separate taxa are inconsistent with this hypothesis.*

Key words: *Endotrypanum* - *Leishmania (Sauroleishmania)* - *Leishmania hertigi* - *Phytomonas* - porcupines - sloths - Phlebotominae - biogeography

In recent years DNA sequence based phylogenies have transformed our understanding of the evolutionary relationships amongst a wide range of protozoa. The growing number of taxa examined and the increasing range of molecules used now makes it possible to consider how host parasite systems have coevolved with more confidence.

The family Trypanosomatidae consists of nine genera of parasitic flagellated protozoa. Five of these genera are parasites of arthropods only and are transmitted contaminatively (Wallace 1979). Three genera have digenetic lifecycles in vertebrates and invertebrates and one genus the *Phytomonas* is parasitic in plants and insects. Phylogenies of the ribosomal RNA genes of members of the Trypanosomatidae indicate that vertebrate parasitism has arisen on at least two separate occasions within this family, once in the *Trypanosoma* and once in the *Leishmania/Endotrypanum* clade (Fernandes et al. 1993, Hollar & Maslov 1997, Lukes et al. 1997, Noyes 1998). The increasing number of taxa used in these phylogenies makes it possible to calibrate these phylogenies against specific events in the evolution of the hosts of these parasites and hence to consider how host parasite systems may have coevolved through time.

The genus *Trypanosoma* is a cosmopolitan parasite of almost all classes of vertebrates and may have very ancient origins in the Palaeozoic. The *Trypanosoma* that are infective to humans may be of much more recent origin in the continents in which they are now found (Stevens et al. 1998). The genus *Phytomonas* is parasitic in flowering plants to which it is transmitted by a range of Hemiptera and Diptera. Since the first flowering plants appeared in the fossil record in the early Cretaceous (130 million years ago - MYA) and began to dominate the terrestrial flora during the second half of the Cretaceous it is possible that the genus *Phytomonas* made the transition from monogenetic parasites of plant feeding insects to digenetic parasites of plants and insects during the Cretaceous. The remaining two genera of digenetic parasites the *Endotrypanum* and the *Leishmania* are the two most closely related genera in the rRNA phylogenies of the Trypanosomatidae and their common ancestor may have made a transition to digenetic parasitism around the time of the mammalian radiation in the late Cretaceous or early Cenozoic (Fernandes et al. 1993).

Leishmania and *Endotrypanum* are both transmitted by phlebotomine sandflies but the genus *Endotrypanum* only infects sloths in the Neotropics whilst the genus *Leishmania* infects at least nine orders of mammals and reptiles and is found in the tropics and subtropics worldwide. Recent phylogenies of the *Leishmania/Endotrypanum* clade have shown that the parasites which are endemic in the New World are closer to the root of this clade than the Old World parasites (Fig. 1). Consequently it has been proposed that this clade made the switch from monogenetic parasites of phlebotomine sandflies to digenetic parasites of sandflies and vertebrates in the Neotropics (Croan et al. 1997, Noyes et al. 1997). This switch is believed to have

Fax: +151-708-8733. E-mail: harry@liv.ac.uk
Received 15 June 1998
Accepted 30 July 1998

occurred at the time of the mammalian radiation at the end of the Cretaceous when the Neotropics were becoming isolated from the rest of the world (Fig. 2). At this time only four groups of mammals are known to have been established in the Neotropics: the monotremes, the marsupials, the notoungulates and the xenarthrans (Fig. 1). The monotremes and notoungulates became extinct in the Neotropics during the Cenozoic leaving the marsupials and xenarthrans as the only indigenous mammals to survive in the Neotropics to the present day (Patterson & Pascual 1972).

The Xenarthra consists of the sloths, the armadillos and the anteaters. The two genera of modern sloths *Bradypus* and *Choloepus* belong to two different families which diverged at the end of the Cretaceous over 70 MYA (Höss et al. 1996). The modern genera are the remnants of a much more diverse group of terrestrial, arboreal and even aquatic sloths that flourished during most of the Cenozoic (Patterson & Pascual 1972, Engelmann 1985, de Muizon & McDonald 1995). Since the two modern genera of sloths are hosts to both genera within the *Leishmania/Endotrypanum* clade it is possible that sloths were the first vertebrate hosts of the common ancestor of these parasites. Monogenetic trypanosomatid parasites of sandflies are not uncommon, although they are rarely described. These trypanosomatids may have become preadapted to development in the blood of vertebrates by frequent exposure to blood in the gut of their sandfly hosts. The relatively low body temperature of modern sloths, and presumably extinct sloths as well, may have facilitated the switch from sandfly to vertebrate hosts. Alternatively it has been suggested that since a number of genera of monogenetic and digenetic trypanosomatids can flourish in the anal scent glands of marsupials, that this organ provided a transitional stage in the acquisition of vertebrate parasitism and that consequently marsupials were the primary hosts of the *Leishmania/Endotrypanum* clade (Deane & Jansen 1988, Jansen et al. 1988). However the sandfly gut is an environment in which the parasite is more likely to be routinely exposed to vertebrate blood over many generations, than in the marsupial anal gland where the parasites have a lower chance of onward transmission. It may be difficult to resolve this question until more is known of the lifecycle and host specificity of the monoxenous trypanosomatids of sandflies.

The phylogenies of the *Leishmania/ Endotrypanum* clade show that the *L. hertigi* complex is more closely related to the genus *Endotrypanum* than to the *Leishmania* (Fig 1.) (Croan et al. 1997, Noyes et al. 1997). Since parasites of the *L. hertigi* complex have only been found in Neotropical porcupines (Rodentia: Hystrichomorpha: Erethizontidae) it is possible that these parasites did not diverge from the *Endotrypanum* line until the hystricomorph rodents arrived in the New World during the Eocene. This provides a calibration point for dating the remaining nodes in the *Endotrypanum/Leishmania* tree (Fig.1). This suggests that the *Endotrypanum* and *Leishmania* genera diverged during the Palaeocene 53-65MYA,

L.(S.)
L.(L.)
L.(V.)
L.h.
E.sp.

CENOZOIC					MESOZOIC			PALAEOZOIC	PRECAMBRIAN
PLIO	MIOCENE	OLI	EOC	PAL	CRETACEOUS	JURAS	TRI		

-5 - Isthmus of Panama - Bering Straits
North Pacific cooling
-26
-37 - Hystricomorpha to Neotropics
-53
-65 - Isolation of Neotropics
-135 - Xenarthra 129 - Phlebotominae
-190 - Marsupalia 173
-225 - Diptera - Mammals / Reptiles
-550 - First vertebrates

Fig. 1: geological periods plotted on a logarithmic scale showing the approximate dates of the appearance of mammal and insect groups important for the understanding of the evolution of the *Endotrypanum/Leishmania* clade. Numbers indicate millions of years before the present. Dates of divergence of mammal groups are from Kumar and Hedges (1998). A phylogeny of the *Endotrypanum/ Leishmania* clade is shown based on the phylogeny of RFLPs of the small subunit ribosomal RNA gene of Noyes et al. (1997). The phylogeny is calibrated against the divergence of the *L. hertigi* and *Endotrypanum* clades which is assumed to have occurred no earlier than the late Eocene at the time the Hystricomorph rodents first appeared in the Neotropical fossil records. *L. (S.)* *L. (Sauroleishmania); L. (L.). L (Leishmania); L. (V.): L. (Viannia); L. h.: L hertigi, E.* sp.. *Endotrypanum* sp.

the Neotropical subgenus *L. (Viannia)* branched off during the early Miocene and the *L. (Leishmania)* and *L. (Sauroleishmania)* diverged during the second half of the Miocene. The tree used to estimate these dates was prepared from RFLPs of the small subunit rRNA gene (Noyes et al. 1997). Further trees including more taxa and data from more genes should increase the reliability of these estimates.

Since the subgenus *L. (Leishmania)* is found in both the Old and New Worlds a member of this subgenus may have migrated to the Old World across the Bering straits region before this region became too cool for the sandfly vectors in the late Miocene (Wolfe 1994). The subgenus *L. (Sauroleishmania)* may then have diverged from the *L. (Leishmania)* in the Old World as a consequence of its adaption to reptiles. Although this hypothesis for the origin of the *L. (Sauroleishmania)* in the Old World requires the minimum number of migrations and extinctions it is not consistent with the phylogenies. If the *L. (Sauroleishmania)* had evolved in the Old World from *L. (Leishmania)* parasites that had migrated from the New World then the RNA and DNA polymerase phylogeny of Croan et al. (1997) would be expected to show the *L. (Sauroleishmania)* branching off between the *L. (L.) mexicana* complex which is restricted to the New World and the *L. (L.) major, L. (L.) tropica* and *L. (L.) donovani* complexes which are restricted to the Old World. Instead it shows that *L. (Sauroleishmania)* branched off before any of the *L. (Leishmania)* subgenus. It is possible that either the *L. (Sauroleishmania)* first evolved in the New World and then also migrated to the Old World independently of the *L. (Leishmania)* before becoming extinct in the New World, or that the common ancestor of both subgenera migrated to the Old World and that the ancestors of *L. mexicana* migrated back after the *L. (Sauroleishmania)* had diverged. However it seems more likely that the position of the *L. (Sauroleishmania)* external to all the *L. (Leishmania)* in the RNA and DNA polymerase phylogeny is an artefact of a faster evolutionary rate in the *L. (Sauroleishmania)*. The long length of the branch leading to the *L. (Sauroleishmania)* is suggestive of a faster evolutionary rate a possibility that was also indicated by a rate test (Croan et al. 1997). Faster evolving groups are known to be pulled towards the outgroup, a phenomenon known as long branch attraction, which could have generated the observed phylogeny (Felsenstein 1988). It is conceivable that the changes that were necessary for the *L. (Sauroleishmania)* to adapt from mammalian hosts to reptile hosts may have forced a temporarily accelerated rate of evolution. It may be possible to test this hypothesis more rigorously by the inclusion of additional taxa in this phylogeny.

It is still not known whether the *L. (Sauroleishmania)* are transmitted by the bite of the sandfly or by contamination when the reptile eats the sandfly (Telford 1995). Since it now appears that the *L. (Sauroleishmania)* have evolved from parasites that are transmitted by the bite of the fly this is perhaps the most likely method for transmission of *L. (Sauroleishmania)* to reptiles as well.

Most *Leishmania* parasites are more restricted in their range of sandfly vectors than in their range of mammalian hosts, implying a much closer coevolutionary relationship with the sandfly than the mammal. However the proposed Neotropical ori-

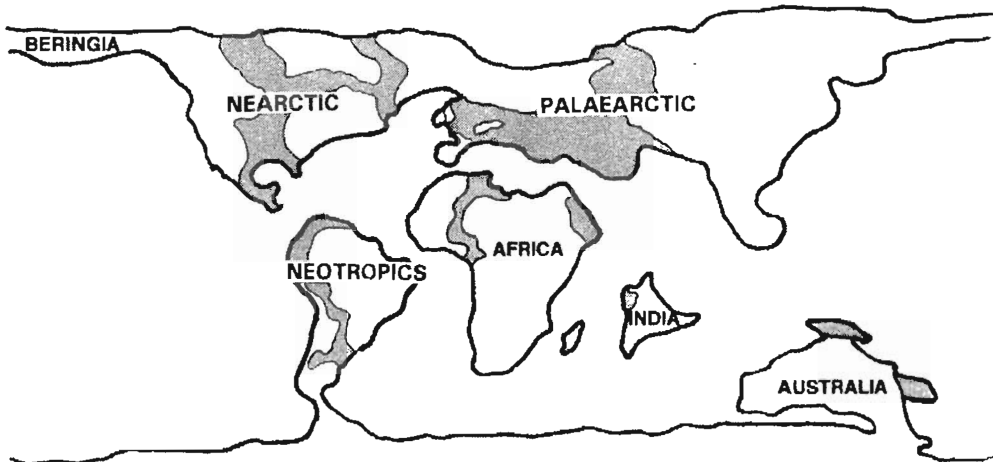


Fig. 2: a map of the world in the late Cretaceous (65-100 million years ago - MYA) showing the isolation of the Neotropics from the Palaearctic. The Isthmus of Panama did not reconnect the two continents until 5 MYA in the Pliocene. The Beringia region which connected the Palaearctic to the Nearctic for most of the Cenozoic (0-65MYA) is shown. The shaded areas indicate shallow seas that covered large parts of the Palaearctic and Nearctic during the Mesozoic and Cenozoic (after Cox 1973).

gin of the *Leishmania* is not consistent with the existing classification of the sandflies. Nevertheless relationships between the subgenera and species complexes of sandflies are still controversial and the existing nomenclature may not reflect the true relationships within this group (Lane 1993). If the hypothesis for a Neotropical origin is correct then it will be possible to make some predictions for the classification of the sandflies that could be tested by molecular methods. In the Old World mammalian *Leishmania* are transmitted by sandflies of the genus *Phlebotomus* and lizard parasites are transmitted by sandflies of the genus *Sergentomyia*. In the New World *Leishmania* and *Endotrypanum* are transmitted by sandflies of the genus *Lutzomyia* (Fig. 3). If *Leishmania* migrated across the Bering region during the Miocene there must have been a resident population of sandfly vectors throughout this area which may have left descendants in both the Old and New Worlds. Consequently the modern sandfly vectors in both the Old and New World may be more closely related to each other than they are to sympatric non-vector genera.

One recent phylogeny of the sandflies does suggest that this is the case. Galati (1995) places the Old World genus *Sergentomyia* in a new subtribe, the Sergentomyiina, with some reptile biting species that are at present in the New World genus *Lutzomyia*. In this classification the Sergentomyiina is clustered in a group of New World subtribes which suggests that the Sergentomyiina may also have evolved in the New World and that the modern *Sergentomyia* are descendants of Sergentomyiina that migrated from the New World to the Old World. This implies that sandflies could have crossed through Beringia at some time. The Sergentomyiina are primarily reptile biters, but members of the *Lutzomyia vexator* series, which Galati places within this subtribe, have been implicated as the vectors of *L. mexicana* in Texas (Kerr et al. 1995). Consequently it is possible that ancestors of the widespread *Lu. vexator* series or of a closely related group may have carried the mammalian *Leishmania* to the Old World. No molecular phylogenies of the Phlebotominae have been published, but they will provide a valuable test of the hypothesis for a Neotropical origin of the *Leishmania*.

REFERENCES

Cox CB 1974. Vertebrate palaeodistributional patterns and continental drift. *J Biogeog 1*: 75-94.

Croan DG, Morrison DA, Ellis JT 1997. Evolution of the genus *Leishmania* revealed by comparison of DNA and RNA polymerase gene sequences. *Mol Biochem Parasitol 89*: 149-159.

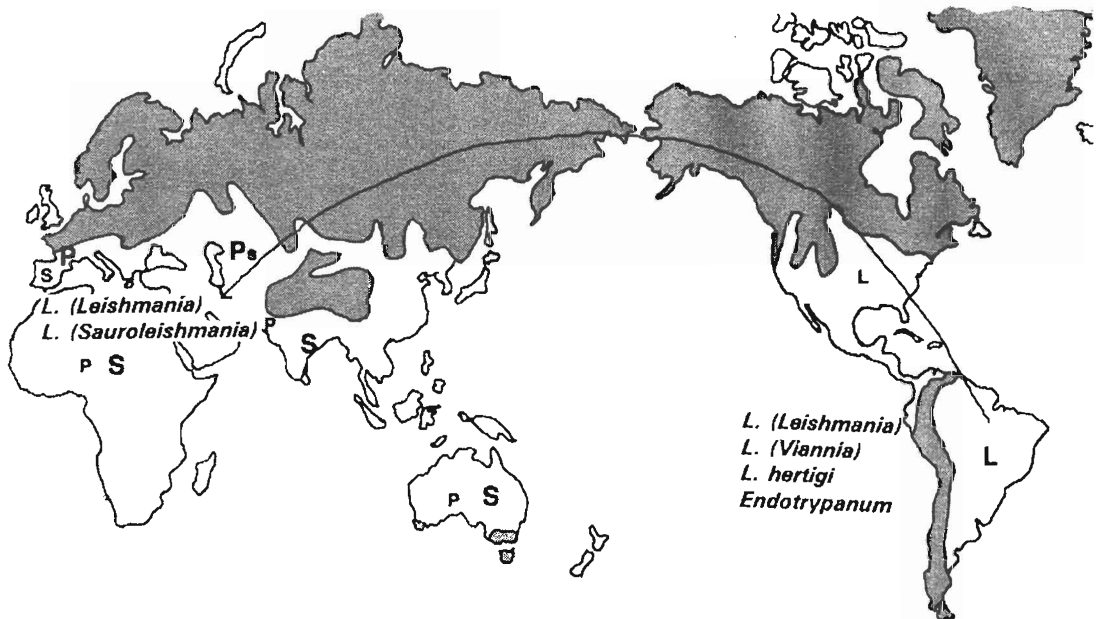


Fig. 3: the modern distribution of sandfly vector genera, after Lewis (1974) and *Leishmania*. The proposed route by which *Leishmania* migrated from the Neotropics to the Old World is indicated by an arrow. The distribution of sandflies is limited by summer temperatures which must remain above 20°C for 50 days a year (Lewis 1982). Areas with a mean temperature of less than 20°C in the hottest month (June for the Northern hemisphere and January for the Southern Hemisphere) are shaded. These areas are not normally suitable for Phlebotominae. In the Northern Hemisphere the 20°C isotherm is at about 45°N, for the Beringia region to be suitable for sandflies this isotherm must have been approximately 15° further north.

de Muizon, C, McDonald, HG 1995. An aquatic sloth from the Pliocene of Peru. *Nature 375*: 224-227

Deane MP, Jansen AM 1988. From a mono to a digenetic life-cycle: how was the jump for flagellates of the family trypanosomatidae. *Mem Inst Oswaldo Cruz 83*: 273-275.

Engelmann GF 1985. The phylogeny of the Xenarthra, p. 51-64. In GG Montgomery, *The Evolution and Ecology of Armadillos, Sloths and Vermilinguas,* Smithsonian Institution Press, Washington.

Felsenstein J 1988. Phylogenies from molecular sequences: inference and reliability. *Annu Rev Genet 22*: 521-565.

Fernandes AP, Nelson K, Beverley SM 1993. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa - perspectives on the age and origins of parasitism. *Proc Natl Acad Sci USA 90*: 11608-11612.

Galati BEA 1995. Phylogenetic systematics of Phlebotominae (Diptera, Psychodidae) with emphasis on American groups. *Boletin de la Direccion de Malariologia y Saneamiento Ambiental 35* (Suppl. 1): 133-142.

Höss M, Dilling A, Currant A, Paabo S 1996. Molecular phylogeny of the extinct ground sloth *Mylodon darwinii. Proc Natl Acad Sci USA 93*: 181-185.

Hollar L, Maslov DA 1997. A phylogenetic view on the genus *Phytomonas*. *Mol Biochem Parasitol 89*: 295-299

Jansen AM, Carreira JC, Deane MP 1988. Infection of a mammal by monogenetic insect trypanosomatids (Kinetoplastida, Trypanosomatidae). *Mem Inst Oswaldo Cruz 83*: 271-272.

Kerr SF, McHugh CP, Dronen Jr NO 1995. Leishmaniasis in Texas: prevalence and seasonal transmission of *Leishmania mexicana* in *Neotoma micropus. Am J Trop Med Hyg 53*: 73-77.

Kumar S, Hedges SB 1998. A molecular timescale for vertebrate evolution. *Nature 392*: 917-920

Lane RP 1993. Sandflies (Phlebotominae), p. 78-119. In RP Lane & RW Crosskey (eds), *Medical Insects and Arachnids*, Chapman & Hall, London.

Lewis DJ 1974. The biology of Phlebotomidae in relation to leishmaniasis. *Ann Rev Entomol 19*: 363-384.

Lewis DJ 1982. A taxonomic review of the genus *Phlebotomus* (Diptera: Psychodidae). *Bull Br Mus Nat Hist (Ent) 45*: 121-209.

Lukes J, Jirku M, Dolezel D, Kral'ová I, Hollar L, Maslov D 1997. Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. *J Mol Evol 44*: 521-527.

Noyes HA 1998. Can Trypanosoma trees be trusted? *Parasitol Today 14*: 49-50.

Noyes HA, Arana BA, Chance ML, Maingon R 1997. The *Leishmania hertigi* (Kinetoplastida; Trypanosomatidae) complex and the lizard *Leishmania*: their classification and evidence for a neotropical origin of the *Leishmania-Endotrypanum* clade. *J Eukaryot Microbiol 44*: 511-517.

Patterson B, Pascual R 1972. The fossil mammal fauna of South America, p. 247-310. In A Keast F Erk & B Glass (eds), *Evolution, Mammals and Southern Continents*, State University of New York Press, Albany.

Stevens JR, Noyes HA, Dover GA, Gibson WC 1998. The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi. Parasitology*, in press.

Telford Jr SR 1995. The kinetoplastid hemoflagellates of reptiles, p. 161-223. In JP Kreier, *Parasitic Protozoa*, 2nd ed., Vol. 10, Academic Press, San Diego.

Wallace FG 1979. Biology of the Kinetoplastida of arthropods, p. 213-240. In WHR Lumsden & DA Evans (eds), *Biology of the Kinetoplastida*, Vol. 2, Academic Press, London.

Wolfe JA 1994. An analysis of Neogene climates in Beringia. *Palaeogeogr Palaeoclimatol Palaeoecol 108*: 207-216.

Genetic Diversity in Natural Populations of New World *Leishmania*

Elisa Cupolillo/+, Hooman Momen*, Gabriel Grimaldi Jr

Laboratório de Leishmaniose, Departamento de Imunologia *Laboratório de Sistemática Bioquímica, Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil

Our results have shown the wide diversity of parasites within New World Leishmania. *Biochemical and molecular characterization of species within the genus has revealed that much of the population heterogeneity has a genetic basis. The source of genetic diversity among* Leishmania *appears to arise from predominantly asexual, clonal reproduction, although occasional bouts of sexual reproduction can not be ruled out. Genetic variation is extensive with some clones widely distributed and others seemingly unique and localized to a particular endemic focus. Epidemiological studies of leishmaniasis has been directed to the ecology and dynamics of transmission of* Leishmania *species/variants, particularly in localized areas. Future research using molecular techniques should aim to identify and follow* Leishmania *types in nature and correlate genetic typing with important clinical characteristics such as virulence, pathogenicity, drug resistance and antigenic variation. The epidemiological significance of such variation not only has important implications for the control of the leishmaniases, but would also help to elucidate the evolutionary biology of the causative agents.*

Key words: New World *Leishmania* - leishmaniasis - epidemiology - molecular characterization

Genetic variation in medically important protozoan parasites and the nature of the reproductive strategies which predispose to such variation are currently the subject of much interest and controversy (Dye et al. 1990, Tibayrenc et al. 1990, 1991, Tibayrenc & Ayala 1991, Hurst et al. 1992, Sibley & Boothroyd 1992). This genetic heterogeneity produces different phenotypes which can be associated with a diversity of clinically important manifestations. At least 13 distinct *Leishmania* species are widespread in the New World and recognized as causing human illness in the Americas. Each of these parasites has a unique zoonotic life cycle, with different sand fly vectors and vertebrate reservoirs (Grimaldi et al. 1989).

Taxonomy of New World Leishmania - Taxonomic studies of leishmanial isolates from the New World indicate tremendous diversity within this genus. A number of new *Leishmania* species from sylvan areas of the Neotropics have been described recently. Some of these parasites are associated with disease in humans, others seem to be restricted to lower orders of mammals, such as rodents and edentates (Lainson & Shaw 1987). Since the original description of these parasites, the number of named species has continually increased and several taxa or classification schemes have been proposed (Gardner 1977, Lainson & Shaw 1979, Rioux et al. 1990), including a subdivision of the genus *Leishmania* Ross, 1903 into two subgenera, the *Viannia* and *Leishmania*, according to the development of the parasite in the gut of sand fly (Lainson & Shaw 1987).

Except for minor differences in size, all species of *Leishmania* are morphologically similar. The initial criteria for identification and classification of these parasites were based on extrinsic characteristic, such as clinical manifestations, geographic and epidemiological features, and a variety of other biologic criteria. However, the variation produced by these criteria lead to the development of biochemical, immunological and molecular methods to provide more precise taxonomic markers based on intrinsic characteristics of the parasites themselves. Among the techniques currently in use are isoenzyme electrophoresis, species-specific monoclonal antibodies or DNA probe and analysis of restriction fragment length polymorphism (RFLP) using different DNA sequences as targets (Macedo et al. 1992, Guizani et al. 1994, Mendonza-Leon et al. 1995).

Multilocus enzyme electrophoresis (MLEE) – The electrophoretic analysis of isoenzymes has been the most widely used method for characterizing *Leishmania* (WHO 1990). Isoenzyme electro-

+Corresponding author. Fax: +55-21-280.1589. E-mail: ecupoli@gene.dbbm.fiocruz.br
Received 15 June 1998
Accepted 30 July 1998

phoresis has the ability to examine a very large sample of structural genes, providing genetic evidence to distinguish polymorphism within species from differences between species as well as information on the reproductive biology of a given organism (Richardson et al. 1986). The technique involves separating isoenzymes by gel electrophoresis and subsequent visualization of specific enzymes using appropriate staining reactions. Isolates with identical banding patterns (alleles) are usually referred to as zymodemes (Godfrey 1979). Important taxonomic information can be obtained by numerical analysis of the electrophoretic bands, which may vary according to the different alleles or genotypic frequencies of loci that are present in distinct parasite strains (Avise 1975).

A large sample of New World *Leishmania* has been analyzed in our laboratories by MLEE. The *Leishmania* strains analyzed until now were grouped in 68 zymodemes (Cupolillo et al. 1994, 1997). Numerical analysis, using phenetic and phylogenetic methods, has demonstrated that the proposed classification of *Leishmania* in two subgenera, *Leishmania* and *Viannia* (Lainson & Shaw 1987), may represent a valid scheme. The parasites were clustered into five phenetic complexes: *L. braziliensis, L. naiffi, L. guyanensis/L. panamensis/L. shawi, L. mexicana, L. major*. All *L. chagasi* parasites formed a unique zymodeme closer to *L. major* than to the other *Leishmania* species. Within the *L. guyanensis/L. panamensis/ L. shawi* complex, we found some named species to be as similar as variant strains within each of these taxa, which showed that these parasites are closely related as a group. The *L. braziliensis* and *L. naiffi* group showed the highest population heterogeneity, presenting 15 and 11 zymodemes, respectively. Some species, like *L. lainsoni, L. equatorensis* and *L. colombiensis* were shown to be very distinct from the other species, but related among themselves (Cupolillo et al. 1994, 1997) (Fig. 1).

Intergenic region typing - DNA analysis provides a means of examining expressed and non-expressed sequences of an organism and is not subject to environmental influences. RFLP analysis detects genetic differences by comparing size variation in DNA banding patterns after restriction endonuclease analysis. The technique has been applied to *Leishmania* focusing basically on the restriction patterns of the minicircle kDNA molecule (Lopes et al. 1984, Pacheco et al. 1986). A PCR based method has been applied in our laboratory to study genetic diversity among parasites, with particular reference to the intraspecific variability that occurs in natural populations of a given *Leishmania* species. In this methodology we amplify the internal transcribed spacers (ITS) of the rRNA gene by PCR, followed by the digestion of the PCR product with several restriction enzymes. The transcribed noncoding regions of rRNA genes (ITS) show extensive variability. Unlike the non-transcribed spacers, the ITS are relatively small (approx. 1 kb in *Leishmania*) and flanked by highly

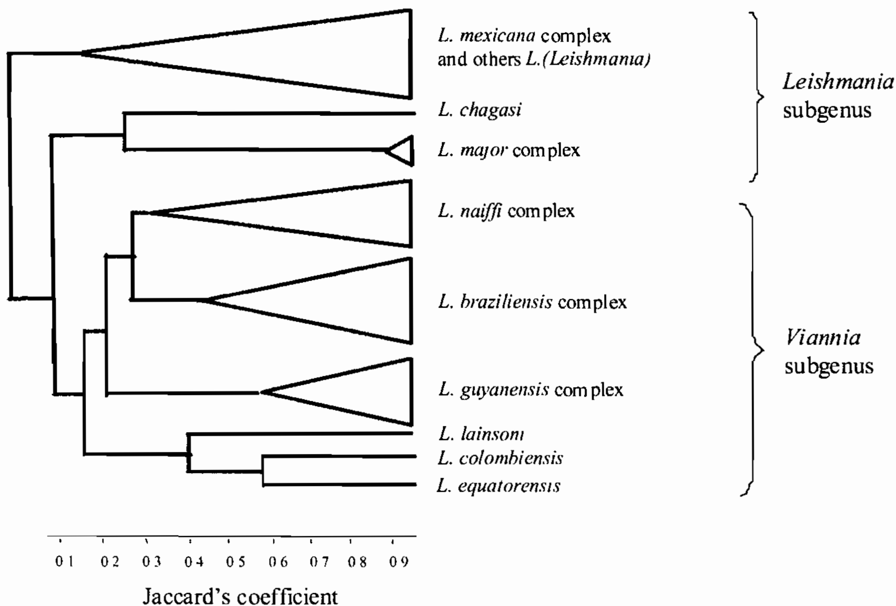


Fig. 1: dendrogram showing the level of similarity and the diversity in each phenetic complex/species.

conserved segments to which PCR primers can be designed. This approach can potentially be applied in many evolutionary situations and amongst a broad range of target loci, we refered to it as intergenic region typing, or IRT (Cupolillo et al. 1995). We analyzed many *Leishmania* isolates representing the *Viannia* subgenus by this method and overall, the ITS clustering showed good agreement with previous organismal or isoenzymatic groupings (Cupolillo et al. 1994). In concordance with the MLEE data *L. braziliensis* and *L. naiffi* population showed a high level of polymorphism. If one accepts current species assignments, one interpretation is that both *L. naiffi* and *L. braziliensis* are highly polymorphic, possibly due to their evolutionary antiquity. This view was supported by cline of evolutionary distances within *L. naiffi*, without clearly predominant zymodeme or ITS subgroups (Cupolillo et al. 1994, 1995). In contrast, the level of polymorphism was less evident in the group *L. guyanensis/L. panamensis/L. shawi*, reinforcing the idea that these species are very closely related (Lainson & Shaw, 1987, Thomaz-Soccol et al. 1993, Cupolillo et al. 1994). However, a small polymorphism was observed at the intra-specific level for *L. guyanensis* and *L. shawi*, in contrast to the results obtained by MLEE (Fig. 2). This result points the IRT as a useful method to study genetic variability amongst intra-species *Leishmania* isolates from an endemic foci.

Genetic variability - The genetic variability was analyzed in the *Viannia* subgenus by population-genetic parameters (Cupolillo et al. 1997). The level of polymorphism was 100%, presenting a media of six alleles/loci. The relative level of genetic variability observed among the parasites indicates that they represent a heterogeneous group

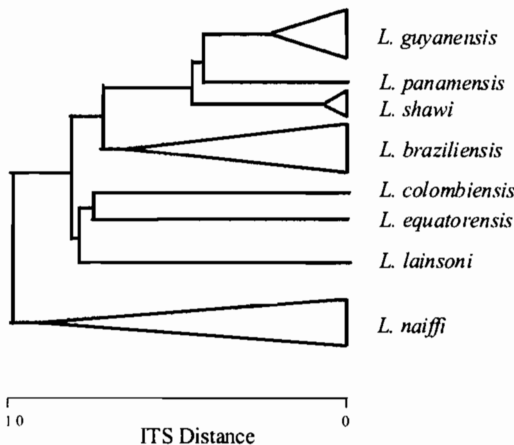


Fig. 2: internal transcribed spacers (ITS) relationship and diversity in *Leishmania (Viannia)* species.

of organisms. The level of heterozygosity observed ($H_L$obs) among the zymodemes was 0.12, whereas the level of heterozygosity expected ($H_L$exp) would be 0.64. If a population is in Hardy-Weinberg equilibrium, both the $H_L$obs and $H_L$exp would be similar. The relatively high values of $H_L$exp and the striking differences found between $H_L$obs and $H_L$exp all pointed to the existence of a clonal structure in natural *Leishmania* populations (Selander & Levin 1980, Tibayrenc & Ayala 1988), reinforced by a strong linkage disequilibrium observed, indicating that asexual reproduction in *Leishmania* is far more common than sexual. Although it is evident that *Leishmania* has a clonal population structure, it does not exclude the possibility of sexual recombination. Our analysis of the MLEE data showed a large number of recurrent mutations in the *Viannia* parasites, which makes it reasonable to attribute some variation to recombination. Moreover, many authors have reported evidence of hybrid formation in *Leishmania* (Evans et al. 1987, Kelly et al. 1991, Darce et al. 1991, Bonfante-Garrido et al. 1992, Dujardin et al. 1993, Belli et al. 1994, Noyes et al. 1996, Bañuls et al. 1997, Delgado et al. 1997), reinforcing the idea that sexual reproduction may occur in *Leishmania*, but at a level as yet undefined. It is important to emphasize that rare or occasional bouts of sexual recombination in a normally asexual organism can have a profound effect on the extent of genetic diversity (Cibulskis 1988).

Transmission cycle - A large number of *Leishmania* isolates have been characterized genetically and considerable variability detected. However, the epidemiolgy of leishmaniasis, in particular the transmission dynamics of the causative agents, is not completely understood. The transmission pattern of New World leishmaniasis involves two distinct cycles: sylvatic and urban. To understand better the role of animal reservoirs and hosts (vertebrate and invertebrate) in these cycles it is important to associated the variability in the parasite populations and the ecology of such endemic areas (Fig. 3).

Characterization by MLEE of *L. chagasi* isolates obtained from a variety of sources (humans, animals and sandflies) have indicated a low level of genetic variation (Momen & Grimaldi 1989), while *L. braziliensis* isolates were highly polymorphic and *L. naiffi* showed intra-specific distances comparable to the largest obtained within all *Viannia* (Cupolillo et al. 1994, 1995, 1997). Interestingly, the same phlebotomine and mammalian species serve as vector and reservoirs of *L. chagasi* throughout its geographic range; with other parasites, such as *L. braziliensis*, several different sand fly and animal species are involved in distinct eco-

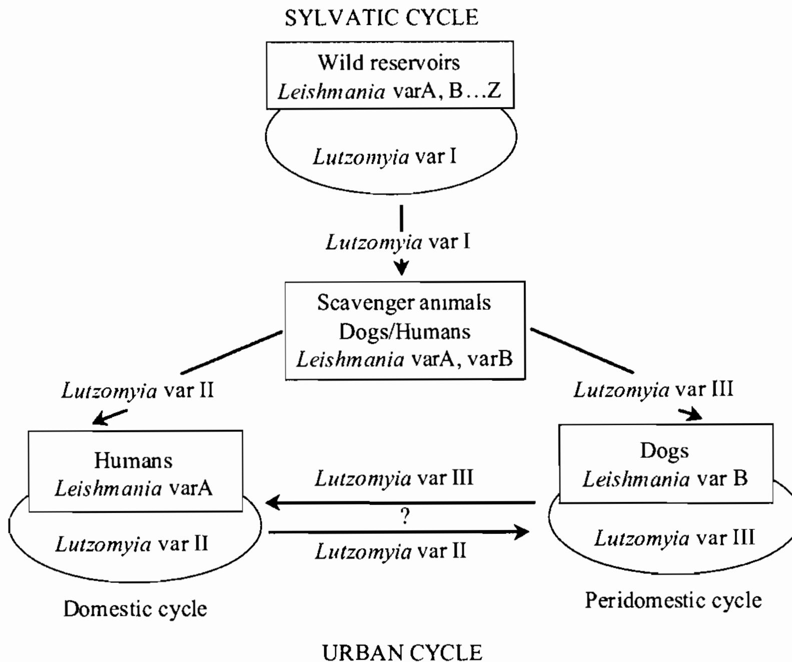


Fig. 3: hypothetical transmissions patterns of *Leishmania*.

logic and geographic regions. Our analyses revealed that the higher molecular diversity found in natural populations of a given *Leishmania* species is related with the higher number of sand fly vector(s) and/or animal reservoir(s) involved in the transmission cycle of the parasites (a co-evolutionary phenomenom?). In contrast, the *L. braziliensis* population circulating in the Brazilian Atlantic coast showed low levels of heterogeneity, and has *Lutzomyia intermedia* as the principal suspected vector. There is no apparent relationship between population heterogeneity in *Leishmania* and the capacity of the parasites to infect their hosts. Accordingly to Mayr (1973), the degree of genetic variability would be comparatively low in those parasites that were naturally selected in single hosts. *Leishmania* parasites may infect several species of host, both vertebrate and invertebrate (Lainson & Shaw 1987, Grimaldi & Tesh 1993). If specificity in the parasite-host relationship is important for *Leishmania* speciation, this process may also be involved in the genetic diversity found among these organisms.

Final comments - The strategies for the prevention and control of leishmaniases are basically the interruption of the transmission cycle (vector and/or reservoir control, personal protection, surveillance, treatment) and vaccination. The control of these diseases is impeded by the lack of vaccination or efficient treatment as well as by the sympatric distribution of *Leishmania* species, the zoonotic nature of the infection and the diversity of the transmission cycles. Although it is not yet evident that the molecular heterogeneity present in *Leishmania* is reflected in properties such as virulence, insect and vertebrate host specificity, geographic range, and drug sensitivity, this polymorphism could have profound consequences for the etiology and treatment of leishmaniases.

REFERENCES

Avise JC 1975. Systematic value of electrophoretic data. *Syst Zool 23:* 465-481.

Bañuls A-L, Guerrini F, Le Pont F, Barrera C, Espinel I, Guderian R, Echeverria R, Tibayrenc M 1997. Evidence for hybridization by multilocus enzyme electrophoresis and random amplified polymorphic DNA between *Leishmania braziliensis* and *Leishmania panamensis/guyanensis* in Ecuador. *J Euk Microbiol 44*: 408-411.

Belli AA, Miles MA, Kelly JM 1994. A putative *Leishmania panamensis/Leishmania braziliensis* hybrid is a causative agent of human cutaneous leishmaniasis in Nicaragua. *Parasitology 109:* 435-442.

Bonfante-Garrido R, Meléndez E, Barroeta S, Mejia de Alejos MA, Momen H, Cupolillo E, McMahon-Pratt D, Grimaldi Jr G 1992. Cutaneous leishmanisis in western Venezuela caused by infection with *Leishmania venezuelensis* and *Leishmania braziliensis* variants. *Trans R Soc Trop Med Hyg 86:* 141-148.

Cibulskis RE 1988. Origins and organization of genetic diversity in natural populations of *Trypanosoma brucei*. *Parasitology 96:* 303-322.

Cupolillo E, Grimaldi Jr G, Momen H 1994. A general classification of New World *Leishmania* using numerical zymotaxonomy. *Am J Trop Med Hyg 50:* 296-311.

Cupolillo E, Grimaldi Jr G, Momen H 1997. Genetic diversity among *Leishmania (Viannia)* parasites. *Ann Trop Med Parasitol 91:* 617-626.

Cupolillo E, Grimaldi Jr G, Momen H, Beverley SM 1995. Intergenic region typing (IRT): a rapid molecular approach to the characterization and evolution of *Leishmania*. *Mol Biochem Parasitol* 73: 145-155.

Darce MJ, Moran X, Palacios A, Belli F, Gomez-Urcuyo F, Zamora D, Valle S, Gantier JC, Momen H, Grimaldi Jr G 1991. Etiology of human cutaneous *Leishmania*sis in Nicaragua. *Trans R Soc Trop Med Hyg 85:* 58-59.

Delgado O, Cupolillo E, Bonfante-Garrido R, Silva S, Belfort E, Grimaldi Jr G, Momen H 1997. Cutaneous leishmaniasis in Venezuela caused by infection with a new hybrid between *Leishmania (Viannia) braziliensis* and *L. (V.) guyanensis*. *Mem Inst Oswaldo Cruz 92*: 581-582.

Dujardin JC, Llanos-Cuentas A, Caceres A, Arana M, Dujardin JP, Guerrini F, Gomez J, Arroyo J, De Doncker S, Jacquet D, Hamers R, Guerra H, Le Ray D, Arevalo J 1993. Molecular karyotype variation in *Leishmania (Viannia) peruviana*: indication of geographical populations in Peru distributed along a north-south cline. *Ann Soc Belge Med Trop 87:* 335-347.

Dye C, Davis CR, Lines JD 1990. When are parasites clonal? *Nature 348:* 120.

Evans DA, Kennedy WPR, Elbihari S, Chapman C, Smith V, Peters W 1987. Hybrid formation in the genus *Leishmania*? *Parassitologia 29:* 165-173.

Gardner PJ 1977. Taxonomy of the genus *Leishmania:* a review of nomenclature and classification. *Trop Dis Bulletin 74:* 1069-1088.

Godfrey DC 1979. The zymodemes of trypanosomes. In Problems in the identification of parasites and their vectors. *Symp Brit Soc Parasitol 17:* 31-53.

Grimaldi Jr G, Tesh RB 1993. *Leishmania*ses of the New World: current concepts and implications for future research. *Clin Microbiol Rev 6:* 230-250.

Grimaldi Jr G, Tesh RB, McMahon-Pratt D 1989. A review of the geographic distribution and epidemiology of *Leishmania*sis in the New World. *Am J Trop Med Hyg 41:* 687-725.

Guizani I, Van Eys GJJM, Ismail RB, Dellagi K 1994. Use of recombinant DNA probes for species identification of Old World *Leishmania* isolates. *Am J Trop Med Hyg 50:* 632-640.

Hurst LD, Hamilton WD, Ladle RJ 1992. Covert sex. *Tree 7*: 144-145.

Kelli JM, Law JM, Chapman CJ, Van Eys GJJM, Evans DA 1991. Evidence of genetic recombination in *Leishmania*. *Mol Biochem Parasitol 46:* 253-264.

Lainson R, Shaw JJ 1979. The role of animals in the epidemiology of the South American Leishmaniasis, p. 1-116. In HR Lumsden, DA Evans (eds). *Biology of the Kinetoplastida*, vol. 2, Academic Press, London.

Lainson R, Shaw JJ 1987. Evolution, classification and geographical distribution, p. 1-120. In W Peters, R Killick-Kendrick, (eds), *The Leishmaniasis, In Biology and Epidemiology*, Vol. 1, Academic Press, London.

Lopes UG, Momen H, Grimaldi Jr G, Marzochi MC, Pacheco RS, Morel CM 1984. Schizodeme and zymodeme characterization of *Leishmania* in the investigation of foci of visceral and cutaneous *Leishmania*sis. *J Parasitol 70:* 89-98.

Macedo AM, Melo MN, Gomes RF, Pena SDJ 1992. DNA fingerprints: a tool for identification and determination of the relationship between species and strains of *Leishmania*. *Mol Biochem Parasitol 53:* 63-70.

Mayr E 1973. The recent historiography of genetics (Essay review). *J Hist Biol 6:* 125-154.

Mendoza-Leon A, Havercroft JC, Barker DC 1995. The RFLP analysis of the β-tubulin gene region in New World *Leishmania*. *Parasitology 111:* 1-9.

Momen H, Grimaldi Jr G 1989. Enzyme electrophoretic evidence for the importation of *L. infantum* into the New World, p. 911-916. In DT Hart, *Leishmaniasis*, Plenum, New York.

Noyes HA, Belli AA, Maingon R 1996. Appraisal of various random amplified polymorphic DNA-polymerase chain reaction primers for *Leishmania* identification. *Am J Trop Med Hyg 55:* 98-105.

Pacheco RS, Lopes UG, Morel CM, Grimaldi GJr, Momen H 1986. Schizodeme analysis of *Leishmania* isolates and comparison with some phenotypic techniques, p. 57-65. In J-A Rioux, *Leishmania. Taxonomy and Phylogeny*, IMEEE, Montpellier.

Richardson BJ, Baverstock PR, Adams M 1986. *Allozyme Electrophoresis*, Academic Press, Sydney, 403 pp.

Rioux J-A, Lanotte G, Serres E, Pratlong F, Bastien P, Perieres J 1990. Taxonomy of *Leishmania* using isoenzymes suggestions for a new classification. *Ann Parasitol Hum Comp 65:* 111-125.

Selander RK, Levin BR 1980. Genetic diversity and structure in *Escherichia coli* populations. *Science 210:* 545-547.

Sibley LD, Boothroyd JC 1992. Virulent strains of *Toxoplasma gondii* comprise a single clonal lineage. *Nature 359:* 82-85.

Thomaz-Soccol V, Lanotte G, Rioux J-A, Pratlong F, Martini-Dumas A, Serres E 1993. Monophyletic origin of the genus *Leishmania* Ross, 1903. *Ann Parasitol Hum Comp 68:* 107-108.

Tibayrenc M, Ayala FJ 1988. Isoenzyme variability in *Trypanosoma cruzi*, the agent of Chagas's disease: genetical, taxonomical, and epidemiological significance. *Evolution 42:* 277-292.

Tibayrenc M, Ayala FJ 1991. Towards a population genetics of micro-organisms: the clonal theory of parasitic protozoa. *Parasitol Today 7:* 228-232.

Tibayrenc M, Kjelberg F, Ayala FJ 1990. A clonal theory of parasitic protozoa: the populations structures of

Entamoeba, Giardia, Leishmania, Naegleria, Plasmodium, Trichomonas and Trypanosoma and their medical and taxonomical consequences. *Proc Natl Acad Sci USA 87:* 2414-2418.

Tibayrenc M, Kjelberg F, Aranaud J, Oury B, Brenière F, Dardé ML, Ayala FJ 1991. Are eukaryotic micro-organism clonal or sexual? A population genetics vantage. *Proc Natl Acad Sci USA 88:* 5129-5133.

WHO-World Health Organization 1990. *Control of the Leishmaniasis*, WHO Tech Rep Ser 793.

The Evolution of Trypanosomes Infecting Humans and Primates

Jamie Stevens/+, Harry Noyes*, Wendy Gibson

School of Biological Sciences, University of Bristol, Bristol BS8 1UG, UK *Liverpool School of Tropical Medicine, Pembroke Place, Liverpool L3 5QA, UK

Based on phylogenetic analysis of 18S rRNA sequences and clade taxon composition, this paper adopts a biogeographical approach to understanding the evolutionary relationships of the human and primate infective trypanosomes, Trypanosoma cruzi, T. brucei, T. rangeli *and* T. cyclops. *Results indicate that these parasites have divergent origins and fundamentally different patterns of evolution.* T. cruzi *is placed in a clade with* T. rangeli *and trypanosomes specific to bats and a kangaroo. The predominantly South American and Australian origins of parasites within this clade suggest an ancient southern super-continent origin for ancestral* T. cruzi, *possibly in marsupials.* T. brucei *clusters exclusively with mammalian, salivarian trypanosomes of African origin, suggesting an evolutionary history confined to Africa, while* T. cyclops, *from an Asian primate appears to have evolved separately and is placed in a clade with* T. (Megatrypanum) *species. Relating clade taxon composition to palaeogeographic evidence, the divergence of* T. brucei *and* T. cruzi *can be dated to the mid-Cretaceous, around 100 million years before present, following the separation of Africa, South America and Euramerica. Such an estimate of divergence time is considerably more recent than those of most previous studies based on molecular clock methods. Perhaps significantly, Salivarian trypanosomes appear, from these data, to be evolving several times faster than* Schizotrypanum *species, a factor which may have contributed to previous anomalous estimates of divergence times.*

Key words: *Trypanosoma brucei* - *Trypanosoma cruzi* - evolution - phylogenetics - small subunit ribosomal RNA - biogeography

The evolutionary relationships of human infective trypanosomes have long been debated (Baker 1963, Hoare 1972, Vickerman 1994). However, it is only recently, with advances in molecular and phylogenetic methods, that significant progress towards resolving trypanosome evolutionary history is being made (e.g. Lake et al. 1988, Fernandes et al. 1993, Wiemer et al. 1995, Maslov et al. 1996, Lukes et al. 1997, Haag et al. 1998, Stevens et al. 1998).

Prior to the advent of molecular techniques, great importance was attached to the mode of transmission as a means of understanding the evolutionary history of trypanosomes (Baker 1963, Hoare 1972). Most vertebrate trypanosomes are transmitted from host to host by bloodsucking arthropod or leech vectors. The trypanosomes are taken up by the vector with a bloodmeal, and usually undergo one or more cycles of development and multiplication in the alimentary tract of the invertebrate, before infective forms are transmitted to a new vertebrate host via saliva, contamination with faeces or ingestion of the whole vector. In this respect the human pathogenic trypanosomes differ markedly. *T. brucei*, the causative agent of African human sleeping sickness, together with a range of related species of veterinary importance (*T. congolense*, *T. simiae* and *T. vivax*), is transmitted by tsetse flies (genus *Glossina*) by the salivarian route. *T. cruzi*, which causes Chagas disease in Latin America, develops in the hindgut of triatomine bugs; infective forms are excreted in the faeces and infect a new host by contamination of wounds or mucous membranes - the stercorarian route (Hoare 1972). The classification and transmission characteristics of another, apparently non-pathogenic species from Latin America, *T. rangeli*, remain much in debate.

In addition to transmission characteristics, *T. brucei* and *T. cruzi* also differ in their mode of infection: *T. brucei* resides in the bloodstream and evades the host immune response by antigenic variation, while *T. cruzi* is an intracellular parasite and multiplies in tissue pseudocysts, with a tran-

This work was supported by the Wellcome Trust (Grant No. 047131/Z/96/Z).
+Corresponding author. Fax: +44-117-925.7374. E-mail: j.r.stevens@bristol.ac.uk
Received 15 June 1998
Accepted 30 July 1998

sient bloodstream phase in the host. It has thus been obvious from even the earliest parasitological studies that *T. brucei* and *T. cruzi* are very different organisms, but, just how different? Such a question has an important bearing on how far results relating to the biochemistry or metabolism of one pathogenic species can be extrapolated to the other, for example, in terms of new chemotherapeutic approaches.

To quantify the evolutionary distance between the two species we have used the divergence of the small subunit ribosomal RNA (ssu rRNA) gene to date the evolutionary split (Stevens et al. 1998). Similar molecular phylogenetic studies have previously relied heavily on "molecular clocks", calibrated by a variety of methods (e.g. Lake et al. 1988, Fernandes et al. 1993, Haag et al. 1998). However, given the almost constant debate surrounding the accuracy of such clocks (Sibley & Ahlquist 1984, Wilson et al. 1987), we have based the date of divergence of *T. brucei* and *T. cruzi* on biogeographical and clade taxon composition (Nelson & Rosen 1981, Meyers & Giller 1988). We believe that this approach is more likely to yield a phylogenetic interpretation with biological relevance, which will contribute to an understanding of the evolution of the genus *Trypanosoma*. Finally, the addition of more taxa, including a primate trypanosome from south-east Asia, *T. cyclops*, has allowed us to explore the robustness of the phylogeny and our biogeographically based evolutionary hypotheses.

MATERIALS AND METHODS

Choice of phylogenetic marker - The 18S ssu rRNA gene was chosen as a suitable phylogenetic marker. It is conserved throughout the eukaryotes, while the range of conserved and variable regions allow diverse rates of genetic evolution to be studied, making it ideal for elucidating both higher evolutionary relationships and those between closely related species (Sogin et al. 1986). Its high copy number also facilitates ease of PCR amplification. The ssu rRNA gene has become the marker of choice for evolutionary analyses of the kinetoplastid protozoa (e.g. Fernandes et al. 1993, Maslov et al. 1994, 1996, Marché et al. 1995, Lukes et al. 1997, Haag et al. 1998).

Trypanosomes - Summary details of all taxa are given in the Table. *T. cyclops* was isolated from *Macaca* spp. in Peninsular Malaysia (Weinman 1972). This uniquely pigmented trypanosome (described as containing large granules of pigment derived from haemoglobin) could not readily be placed in any existing subgenus. The vector is unknown, but transmission by reduviid bugs was ruled out (Weinman 1972).

Ribosomal RNA sequences - The ssu rRNA sequence of *T. cyclops* was sequenced, as described by Stevens et al. (1998). Briefly, the gene was amplified by PCR from trypanosome template DNA as a fragment of ~2 kb using conserved primers (Maslov et al. 1996). The products of 8-10 separate PCR reactions were then purified and pooled, prior to automated sequencing in both directions at approximately 300 base pair intervals using 12 additional internal primers (Maslov et al. 1996) on a Perkin-Elmer ABI 377 automated sequencer. A consensus sequence was assembled for each trypanosome strain from the internal primer sequences using AutoAssembler v.2.0 (Applied Biosystems, Perkin-Elmer).

Thirty-two *Trypanosoma* sp. sequences were included from Stevens et al. (1998; Table), together with 15 *Trypanosoma* sp. sequences from the EMBL/GenBank databases. The suitability of free-living bodonid taxa as outgroups for phylogenetic analysis of trypanosomatids has been established by a number of studies using a range of ribosomal and protein coding genes (see Stevens et al. 1998). In this study, *Trypanosoma* species were compared with a range of ten outgroup taxa (*Bodo caudatus* X53910; *Trypanoplasma borreli* L14840; *Crithidia* spp. X03450, L29264; *Leishmania* spp. X53912, X07773, X53913, X53915; and *Phytomonas* spp. L35076, L35077).

Alignments - The *T. cyclops* sequence was incorporated into the alignment of Stevens et al. (1998). In this alignment, all sequences were aligned primarily to eight *Trypanosoma* sequences downloaded from the rRNA database maintained at the University of Antwerp (Neefs et al. 1990); the alignment of these eight template sequences is based on their secondary structure. Sub-sections of the alignment, between 'anchor' regions of high homology were then sub-aligned using the program Clustal V (Higgins et al. 1992), before final adjustments were made by eye. Hypervariable sites, where nucleotide changes were saturated, and regions where it was not possible to produce a single reliable alignment across all 58 taxa were excluded from the analysis. Following this, a number of separate alignments, representing more or less stringent subsets of a 'standard' alignment, were explored (Stevens et al. 1998) and used as the basis for the phylogenetic analysis presented in this paper. Certain sites which were locally informative between closely related taxa introduced 'noise', resulting in a loss of definition (reduced bootstrap support) at higher phylogenetic levels; such sites were excluded from the final analysis (Fig.) and the alignment used included 1801 nucleotide positions (available on request from JRS).

Phylogenetic analyses - Bootstrapped maxi-

TABLE

Summary details of *Trypanosoma* spp. analysed[a]

Species	Host		Location
T. avium	Bird	*Fringilla coelebs*	Czech Republic
T. brucei gambiense	Human	*Homo sapiens*	Nigeria
T brucei rhodesiense	Human	*Homo sapiens*	Uganda
T. cobitis	Freshwater fish	*Noemacheilus barbatulus*	England
T. congolense (kilifi)	Domestic goat	*Capra* sp.	Kenya
T. congolense (forest	Domestic goat	*Capra* sp.	Cameroon
T. congolense (savannah)	Domestic goat	*Capra* sp.	Kenya
T. cruzi (Z I)	Human	*Homo sapiens*	Brazil
T. cruzi (Z II)	Triatomine bug	*Triatoma infestans*	Chile
T. cruzi (Z III)	Human	*Homo sapiens*	Brazil
T. cruzi marinkellei	Bat	*Phyllostomum discolor*	Brazil
T. cyclops	Macaque	*Macaca* sp.	Malaysia
T. dionisii	Bat	*Pipistrellus pipistrellus*	England
T. dionisii	Bat	*Pipistrellus pipistrellus*	Belgium
T. equiperdum	Horse	*Equus caballus*	China
T. evansi	Capybara	*H. hydrochaeris*	Brazil
T. godfreyi	Tsetse fly	*G.m.submorsitans*	The Gambia
T. grayi	Tsetse fly	*G.p.gambiensis*	The Gambia
T. lewisi	Rat	*Rattus* sp.	England
T. mega	Toad	*Bufo regularis*	Africa
T. microti	Vole	*Microtis agrestis*	England
T. pestanai	Badger	*Meles meles*	France
T. rangeli	Dog	*Canis* sp.	Venezuela
T. rotatorium	Frog	*Rana catesbeiana*	Canada
T. simiae	Tsetse fly	*G.m.submorsitans*	The Gambia
T. theileri	Cattle	*Bos taurus*	Germany
T. theileri	Cattle	*Bos taurus*	Scotland
T. varani	Lizard	*Varanus exanthematicus*	Senegal
T. vespertilionis	Bat	*Pipistrellus pipistrellus*	England
T. sp.	Leech	*Piscicola geometra*	England
T. sp.	Kangaroo	*Macropus giganteus*	Australia
T. sp.	Wombat	*Vombatus ursinus*	Australia
T. sp.	Deer	*Cervus dama*	Germany

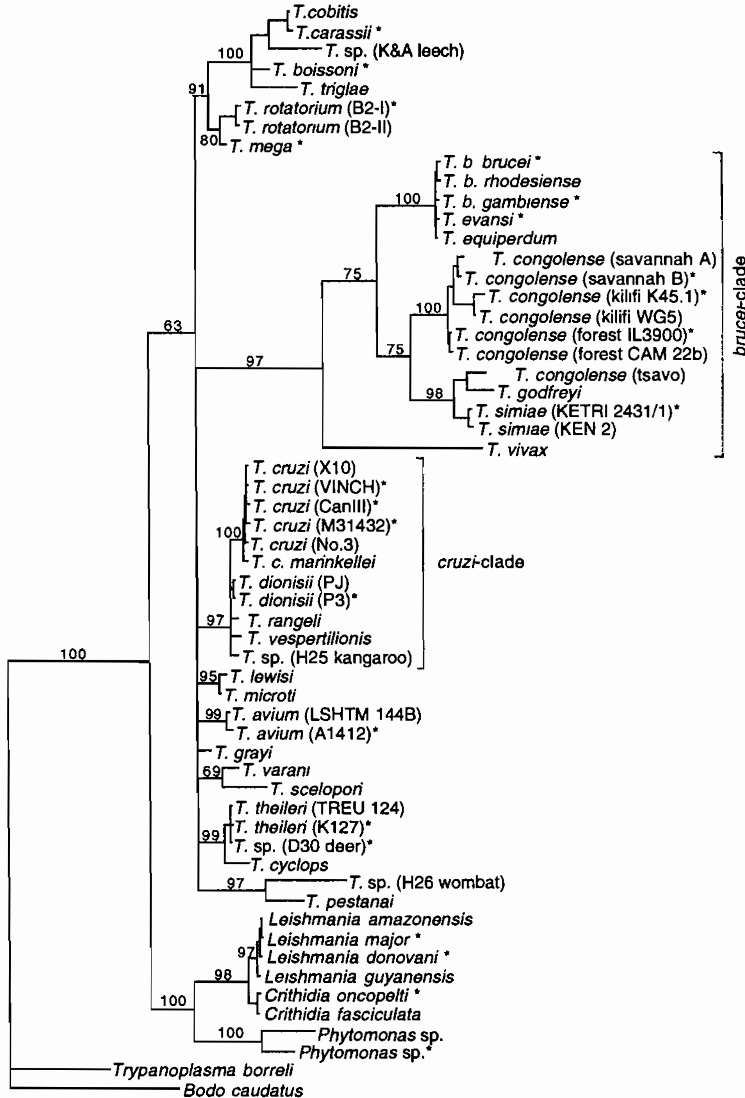
a: details of the 33 *Trypanosoma* sp. isolates sequenced as part of this study. See Stevens et al. (1998) for additional isolation details; see Weinman (1972) for full details of *T. cyclops*. Information on the 15 additional *Trypanosoma* sp. ssu rRNA sequences added from EMBL/GenBank can be obtained from database: *T. boissoni* U39580; *T. carassii* L14841; *T. rotatorium* U39583; *T. triglae* U39584; *T. brucei brucei* M12676; *T. congolense* (kilifi-type) U22317; *T. congolense* (forest-type) U22319; *T. congolense* (savannah-type) U22315; *T. congolense* (tsavo-type) U22318; *T. simiae* U22320; *T. vivax* U22316; *T. cruzi* X53917; *T. cruzi* M31432; *T. avium* U39578; *T. scelopori* U67182.

mum parsimony analysis of 58 kinetoplastid 18S ssu rRNA sequences (Table) was performed with 100 replicates (Fig.); again, analyses were repeated with a number of more and less stringent alignments (see Stevens et al. 1998). The number of taxa necessitated the use of a heuristic search strategy to find the most parsimonious trees. The default options of PAUP were used: TBR branch swapping, zero length branches collapsed and 10 random addition sequences (bootstrap analyses used simple addition).

Maximum-likelihood analysis was also performed; however, due to computational constraints, a reduced data set of 36 taxa was employed. Taxa were selected for inclusion in the maximum likelihood analysis so as to maximise the degree of sequence variation analysed, highly homologous sequences being excluded; starting trees were derived by both parsimony and neighbour-joining. Transition/transversion ratios were estimated from the data in preliminary runs and then set for full analyses. All analyses were performed using test version 4.0d63 of PAUP*, written by David L Swofford.

RESULTS

Phylogenetic analysis. The phylogram (Fig.) classifies the genus *Trypanosoma* into three major clades, including the *brucei*-clade and the *cruzi*-



Phylogram constructed by bootstrapped (100 replicates) maximum parsimony analysis of 58 kinetoplastid 18S ssu rRNA sequences and rooted on the free-living kinetoplastid, *Bodo caudatus*. The tree is derived from the 28 most parsimonious trees of length = 1135 (RI = 0.8177, CI = 0.5570), based an alignment of 1801 nucleotide sites. Bootstrap values for all major nodes are given and all branches receiving bootstrap support values >50% are shown; relationships failing to achieve this level of support are shown as polytomies (i.e. branch points at which three or more branches arise from the ancestral line). Certain clades, referred to in the text, are defined by dashed brackets. The 22 taxa not included in the maximum-likelihood analyses are marked *. Details of all taxa are given in the Table.

clade [the *brucei* and *cruzi* clades (and all species therein) are defined in the Fig. The terms '*brucei*-clade' and '*cruzi*-clade' are used throughout this study to refer to the clades containing (a) all mammalian Salivarian (Hoare 1972) trypanosomes (*brucei*-clade) and (b) trypanosomes in the subgenus *Schizotrypanum*, plus *T. rangeli* and an as yet unidentified species of trypanosome from a kangaroo (*cruzi*-clade)], and six minor clades, which together form a nine-way polytomy. Importantly, the phylogram is proven to be robust and its structure is largely the same as that presented by (Stevens et al. 1998). However, the addition of two *Phytomonas* sp. to the analysis significantly reduces phylogenetic definition at the upper level of the *Trypanosoma*, such that the clade containing trypanosomes from aquatic and related hosts no longer diverges earlier than other *Trypanosoma*, and forms part of a nine-way polytomy within the genus. Such a result serves to underline the im-

portant influence that the choice of outgroup taxa may exert on phylogenetic analyses.

T. cyclops, isolated from a Malaysian primate, is classified with *T. theileri* and another well characterized *Megatrypanum* species from a deer (Fig.; 100% bootstrap support), rather than in either the *brucei*-clade or the *cruzi*-clade, which contain all other human/primate infective trypanosomes [Johnson (1933) reported *T. lewisi* in the blood of a child, also in Malaysia]. Such a result does not support a close phylogenetic relationship with either African or South American human/primate infective trypanosomes, whilst its apparent lack of overt *Megatrypanum* characteristics (Weinman 1972) calls into question the taxonomic basis of the subgenus *Megatrypanum*.

The results of the parsimony analyses were again strongly supported by maximum-likelihood analysis. The positions and branching order of all major clades were identical between methods (irrespective of starting tree), and only minor variations in the positions of certain terminal taxa were apparent. Indeed, the main phylogenetic relationships revealed in the tree are as described by Stevens et al. (1998) and are largely robust to the addition of the primate trypanosome and additional outgroup species.

Briefly, the phylogenetic analysis confirms the monophyly of the genus *Trypanosoma* with bootstrap support of 63% (Fig.). The human pathogenic trypanosomes, *T. brucei* and *T. cruzi*, are placed in separate clades, each receiving high bootstrap support of >97%. The *brucei*-clade, contains all species of mammalian Salivarian trypanosomes (Hoare 1972). Except for *T. evansi* and *T. equiperdum*, these trypanosomes are all of African origin and transmitted by tsetse flies [analysis of kinetoplast (mitochondrial) DNA (Borst et al. 1987) and isoenzymes (Lun et al. 1992, Gibson et al. 1983) points to *T. evansi* and *T. equiperdum* being comparatively recent mutants of *T. brucei*, which have been able to spread outside Africa because they no longer rely on tsetse transmission; the particularities of these two species are therefore irrelevant to the more ancient evolution of the clade]. The host exclusivity of this clade suggests a distinct evolutionary history initially confined to Africa. Trypanosomes of African origin from other vertebrates are completely unrelated (e.g. *T. grayi*, *T. varani* from African reptiles; *T. mega* from an African toad). A similar result is reported by Haag et al. (1998).

The *cruzi*-clade contains all subgenus *Schizotrypanum* species - *T. cruzi* isolates from humans, sylvatic and domestic mammals, including bats and opossums, together with trypanosomes specific to Old and New World bats, *T. rangeli* and an as yet unidentified trypanosome species from an Australian kangaroo. The origins of parasites within this clade thus lie largely in South America and Australia; the only trypanosomes from this clade representing the Old World are those infecting bats.

The taxonomic and evolutionary status of human infective *T. rangeli* (generally classified as subgenus *Herpetosoma*) remains controversial (D'Alessandro & Saravia 1992, Stevens & Gibson 1998). In the current study *T. rangeli*, albeit only a single isolate (RGB - Basel), is classified firmly in a clade with a range of *Schizotrypanum* species (bootstrap 97%); the classification of this isolate as *T. rangeli* is supported by preliminary results from analysis of the miniexon which indicate it to be of the correct size and sequence according to Murthy et al. (1992). *T. rangeli* and *T. cruzi* also cluster together (bootstrap >90%) and separate from Salivarian trypanosomes in phylogenetic analyses of miniexon sequences (Stevens & Gibson, unpublished data).

Phylogenetic resolution - Despite support for at least nine distinct clades within the *Trypanosoma*, it is not possible on the basis of these rRNA data to determine the exact order in which these clades diverged and a nine-way polytomy within the genus remains unresolved (Fig.). This may be due to limitations on the resolution of the ssu gene over this time scale, to a possibly explosive divergence of trypanosome species over a very short period some time around 100 million years before present (mybp) or, as seems probable, to a combination of both these factors.

Nevertheless, the inclusion of a large and varied range of taxa (Swofford et al. 1996) has enabled elucidation of the complex relationships of the human infective trypanosomes and, while saturation of some variable regions within the ssu gene may preclude the accurate determination of branching order at more ancient levels, considerable support for the 'correctness' of the phylogenetic relationships represented in the tree is provided by the logical placement of the outgroup trypanosomatids, *Leishmania* spp., *Phytomonas* spp. and *Crithidia* spp., which are well separated from the *Trypanosoma* (Fig.).

Comparative rates of sequence evolution - The phylogenetic analysis provides evidence of very different rates of evolution within the clades containing *T. brucei* and *T. cruzi*. Comparison of deeper branch lengths suggests a difference in intra-clade evolution rate of approximately 8-fold. The exact extent to which the rapid evolution of certain lineages within the Salivarian clade may have distorted the topology of the tree (and hence estimates of evolutionary rates) is unknown. Nevertheless, the tree appears sufficiently robust to have avoided the Salivaria being drawn towards

outgroup taxa by long-branch attraction (Felsenstein 1978, Hendy & Penny 1989), a problem encountered in many previous studies (e.g. Fernandes et al. 1993, Maslov et al. 1994, 1996).

DISCUSSION

Phylogenetic analysis of variation in ssu rRNA genes of the 48 trypanosome specimens places the two human pathogens, *T. brucei* and *T. cruzi*, unequivocally in two distinct clades. The Asian primate trypanosome, *T. cyclops*, is grouped in an apparently unrelated clade with trypanosomes from cattle and deer.

The time of divergence of *T. brucei* and *T. cruzi* can be estimated by a range of methods including: sequence divergence analysis (the molecular clock approach), by reference to host phylogenies and by consideration of palaeo- and biogeographical data.

The concept of a molecular clock was first proposed by Zuckerkandl and Pauling (1965). Since then the exact nature of the workings of such 'clocks' have remained under almost constant debate (Sibley & Ahlquist 1984, Wilson et al. 1987). It is apparent that, if they do exist, they are at best only stochastically constant (Fitch 1976), and that different types of DNA sequence undoubtedly evolve at significantly different rates. Nevertheless, within given taxonomic groups and defined categories of genetic marker, the concept of a molecular clock can provide a useful tool for understanding phylogenetic relationships.

In the study of trypanosome evolution molecular clocks have been used by a number of authors to attempt to date the divergence events between important taxonomic groups. In a recent and comprehensive study, Haag et al. (1998) used an estimate of 0.85% substitution per 100 million years, derived from rRNA based studies of Apicomplexa (Escalante & Ayala 1995), to date the divergence of Salivarian trypanosomes from other trypanosomes at about 300 mybp. Such a result places the divergence of *T. brucei* and the Salivaria in the late Carboniferous, at a time when the very first reptiles had just appeared in a world dominated by amphibians, and long before the appearance of even the most primitive mammals. Perhaps significantly, the continents with which several major extant groups of trypanosomes (i.e. Africa: Salivaria; South America: *Schizotrypanum*) are now generally associated, had not at that time even begun to separate, but were grouped together in the solid southern land mass known as Gondwana (Cox & Moore 1993, Smith et al. 1994).

A second method for estimating times of divergence in parasite phylogenies is based on congruence of host and parasite phylogenies, a much debated concept. Using this approach, parasite trees can be calibrated by reference to time points within host phylogenies, which have been constructed on the basis of independent evidence, e.g. fossils. Such an approach was also used by Haag et al. (1998), who used the divergence of fish from higher vertebrates (400 mybp) and the divergence of birds from rodents (220 mybp), to estimate the split of Salivarian trypanosomes from other trypanosomes at 260 and 500 mybp, respectively. Again, even the most recent of these estimates places the divergence of the Salivaria somewhat unrealistically in the mid-Permian.

A third approach to calibrating organismal phylogenies is by reference to known biogeographical events - vicariance biogeography (Nelson & Rosen 1981) - and several sequence divergence based studies of trypanosomatids have drawn on this technique, for example, to date the divergence of *Leishmania* and *Trypanosoma* (Lake et al. 1988) and to date the split between Old and New World *Leishmania* (Nelson et al. 1990, Fernandes et al. 1993). Using this approach we previously obtained a mid-Cretaceous date for the divergence of *T. brucei* and *T. cruzi* (Stevens et al. 1998). In summary, we suggested that the exclusively African mammalian tsetse-transmitted taxon complement of the *brucei*-clade (excepting *T. evansi* and *T. equiperdum* - see above), points to an origin in Africa. The first time at which Africa became isolated was around 100 mybp, in the mid-Cretaceous, when it finally separated from South America and Euramerica (Parrish 1993, Smith et al. 1994). At this time, the first mammals were present, but had not yet begun major diversification and it is easy to envisage subsequent coevolution of this clade with ancient African hosts.

The *cruzi*-clade would thus have a southern super-continent (South American/Antarctica/Australia) origin, an interpretation which makes sense of the inclusion of the Australian marsupial trypanosome in the clade. Indeed, the early evolution of this clade may have been associated with the dominant marsupial fauna of the region. The opossum, *Didelphis* sp., a not so distant relative of the Australian kangaroos (Flannery 1989), is a particularly important natural reservoir of *T. cruzi* in South America and can maintain a patent parasitaemia throughout its life, with no apparent clinical symptoms (Deane et al. 1986). The only trypanosomes from this clade found in the Old World are those infecting bats. The biological similarity of *T. cruzi* and certain species of bat trypanosomes has been recognised for some time - all are classified in the subgenus *Schizotrypanum*. The present day distribution can be explained by the ability of bats to disperse over long distances, particularly across

water barriers and, while *T.* (*Schizotrypanum*) species have been isolated from European bats on a number of occasions (Baker 1974, Baker & Thompson 1971, Baker & Mewis 1987), reports of *T. cruzi*-like trypanosomes from other hosts in the Old World are insubstantial (Hoare 1972).

In the current study, *T. rangeli* was placed firmly within the *cruzi*-clade. However, the taxonomic position of this human infective species has long been disputed. While it appears morphologically and behaviourally similar to subgenus *Herpetosoma* trypanosomes (Hoare 1972), it is transmitted by both salivarian and stercorarian routes (D'Alessandro & Saravia 1992), while Añez (1982) even separated it into a new subgenus, *Tejeraia*. A limited study based on ß-tubulin gene sequences (Amorim et al. 1993) suggested *T. rangeli* to be more closely related to *T. brucei* than to *T. cruzi*. While this result cannot be disputed, it is widely recognised (Swofford *et* al. 1996) that studies including limited numbers of taxa spanning disparate levels of relatedness are highly prone to artifactual effects. Certainly, the close relationship between *T. rangeli* and *T. cruzi* evident from ssu rRNA analysis has also been confirmed by comparison of miniexon sequences (Stevens & Gibson, unpublished data). The miniexon sequence also confirms that the *T. rangeli* isolate used in the current study is a *bona fide T. rangeli* (Murthy et al. 1992).

The classification of *T. cyclops* with otherwise apparently unrelated *T.* (*Megatrypanum*) species and apart from other human/primate trypanosomes, suggests that its ability to infect primates has evolved independently (presumably in Asia) from species in either of the two clades containing human infective trypanosomes.

From the separate evolutionary histories of *T. brucei* and *T. cruzi* constructed from the phylogenetic evidence, we can deduce that their pathogenicity to humans developed on very different time scales. In Africa, *T. brucei* would have effectively co-evolved with hominids, since the first hominids evolved 5-15 mybp, the genus *Homo* 3 mybp (Johanson & Taieb 1993) and *Homo sapiens* not earlier than 300 000 years bp, presumably in continuous contact with both trypanosomes and tsetse flies. In contrast, human contact with *T. cruzi* would not have occurred prior to human migration into the Americas, which is generally dated no earlier than 30 - 40 000 years bp. Moreover, there is no evidence for contact earlier than 3000 years bp when the first permanent settlements were made by previously nomadic cultures (Rothhammer et al. 1985). Humans, like other primates would have become infected as a simple addition to the already extensive host ranges of *T. cruzi* and *T. rangeli* (Hoare 1972).

Finally, to what extent their different evolutionary histories have affected intra-clade evolution rates is unknown, however, it appears that *brucei*-clade species are diverging at a rate up to eight times faster than that observed amongst *cruzi*-clade species. Moreover, such differences in evolutionary rates between trypanosome clades are in keeping with results from previous studies (Maslov et al. 1996) and are confirmed by maximum-likelihood based rate analyses (Stevens, Rambaut & Gibson, unpublished data), which indicate the differences in rates of evolution between the Salivarian clade and the *Schizotrypanum* clade to be significant. Reasons for these rate differences remain to be explored.

ACKNOWLEDGEMENTS

To Chris Schofield for valuable discussions on the evolutionary relationships of *T. cruzi*.

REFERENCES

Amorim MI, Momen H, Traub-Cseko YM 1993. *Trypanosoma rangeli*: sequence analysis of ß-tubulin gene suggests closer relationship to *Trypanosoma brucei* than to *Trypanosoma cruzi*. *Acta Trop 53*: 99-105.

Añez N 1982. Studies on *Trypanosoma rangeli* Tejera, 1920. IV. A reconsideration of its systematic position. *Mem Inst Oswaldo Cruz 77*: 405-415.

Baker JR 1963. Speculations on the evolution of the family Trypanosomatidae Doflein, 1901. *Exp Parasitol 13*: 219-233.

Baker JR 1974. Protozoan parasites of the blood of British wild birds and mammals. *J Zool 172*: 169-190.

Baker JR, Mewis GE 1987. *Trypanosoma* (*Schizotrypanum*) sp. indet. from a Maltese bat. *Acta Trop 44*: 99-100.

Baker JR, Thompson GB 1971. Two species of *Trypanosoma* from British bats. *Trans R Soc Trop Med Hyg 65*: 427.

Borst P, Fase-Fowler F, Gibson WC 1987. Kinetoplast DNA of *Trypanosoma evansi*. *Mol Biochem Parasitol 23*: 31-38.

Cox CB, Moore PD 1993. *Biogeography: An Ecological and Evolutionary Approach,* 5th ed, Blackwell, Oxford, 326 pp.

D'Alessandro A, Saravia NG 1992. *Trypanosoma rangeli*, p. 1-54. In JP Kreier, JR Baker (eds), *Parasitic Protozoa*, 2nd ed, vol. 2, Academic Press, San Diego.

Deane MP, Lenzi HL, Jansen AM 1986. Double development cycle of *Trypanosoma cruzi* in the opossum. *Parasitol Today 2*: 146-147.

Escalante AA, Ayala FJ 1995. Evolutionary origin of *Plasmodium* and other Apicomplexa based on rRNA genes. *Proc Natl Acad Sci USA 92*: 5793-5797.

Felsenstein J 1978. Cases in which parsimony or compatibility will be positively misleading. *Syst Zool 27*: 401-410.

Fernandes AP, Nelson K, Beverley SM 1993. Evolution of nuclear ribosomal RNAs in kinetoplastid proto-

zoa: Perspectives on the age and origins of parasitism. *Proc Natl Acad Sci USA 90*: 11608-11612.

Fitch WM 1976. Molecular evolutionary clocks, p. 160-178. In FJ Ayala, *Molecular Evolution*, Sinauer, Sunderland, MA.

Flannery TF 1989. Phylogeny of the Macropodoidea; a study in convergence, p. 1-46. In G Grigg, P Jarman, I Hume (eds), *Kangaroos, Wallabies and Rat-kangaroos*, vol. 1, Surrey Beatty & Sons, Chipping Norton, NSW.

Gibson WC, Wilson AJ, Moloo SK 1983. Characterisation of *Trypanosoma* (*Trypanozoon*) *evansi* from camels in Kenya using isoenzyme electrophoresis. *Res Vet Sci 34*: 114-118.

Haag J, O'Huigin C, Overath P 1998. The molecular phylogeny of trypanosomes: evidence for an early divergence of the Salivaria. *Mol Biochem Parasitol 91*: 37-49.

Hendy MD, Penny D 1989. A framework for the quantitative study of evolutionary trees. *Syst Zool 38*: 297-309.

Higgins DG, Bleasby AJ, Fuchs R 1992. Clustal V: improved software for multiple sequence alignment. *Comp Applns Biosci 8*: 189-191.

Hoare CA 1972. *The Trypanosomes of Mammals*, Blackwell, Oxford, 749 pp.

Johanson DC, Taieb M 1976. Plio-Pleistocene hominid discoveries in Hadar, Ethiopia. *Nature 260*: 293-297.

Johnson PD 1933. A case of infection by *Trypanosoma lewisi* in a child. *Trans R Soc Trop Med Hyg 26*: 467.

Lake JA, De La Cruz VF, Ferreira PCG, Morel C, Simpson L. 1988. Evolution of parasitism: Kinetoplastid protozoan history reconstructed from mitochondrial rRNA gene sequences. *Proc Natl Acad Sci USA 85*: 4779-4783.

Lukes J, Jirku M, Dolezel D, Kral'ova I, Hollar L, Maslov DA 1997. Analysis of ribosomal RNA genes suggests that trypanosomes are monophyletic. *J Mol Evol 44*: 521-527.

Lun Z-R, Allingham R, Brun R, Lanham S 1992. The isoenzyme characteristics of *Trypanosoma evansi* and *Trypanosoma equiperdum* isolated from domestic stocks in China. *Ann Trop Med Parasitol 86*: 333-340.

Marché S, Roth C, Philippe H, Dollet M, Baltz T 1995. Characterization and detection of plant trypanosomatids by sequence analysis of the small subunit ribosomal RNA gene. *Mol Biochem Parasitol 71*: 15-26.

Maslov DA, Avila HA, Lake JA, Simpson L 1994. Evolution of RNA editing in kinetoplastid protozoa. *Nature 368*: 345-348.

Maslov DA, Lukes J, Jirku M, Simpson L 1996. Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. *Mol Biochem Parasitol 75*: 197-205.

Meyers AA, Giller PS 1988. *Analytical Biogeography*, Chapman & Hall, London, 578 pp.

Murthy VK, Dibbern KM, Campbell DA 1992. PCR amplification of mini-exon genes differentiates *Trypanosoma cruzi* from *T. rangeli*. *Mol Cellular Probes 6*: 237-243.

Neefs J-M, Van De Peer Y, Hendriks L, De Wachter R 1990. Compilation of small ribosomal subunit RNA sequences. *Nucleic Acids Res 18*: 2237-2243.

Nelson G, Rosen DE 1981. *Vicariance Biogeography: A Critique*, Columbia University Press, New York, 500 pp.

Nelson K, Alonso G, Langer PJ, Beverley SM 1990. Sequence of the dihydrofolate reductase-thymidylate synthase (DHFR-TS) gene of *Leishmania amazonensis*. *Nucleic Acids Res 18*: 2819.

Parrish JT 1993. The palaeogeography of the opening South Atlantic, p. 8-27. In W George, R Lavocat (eds), *The Africa-South America Connection*, Oxford Monographs on Biogeography No.7, Clarendon Press, Oxford.

Rothhammer F, Allison MI, Nunez L, Standen V, Arriaza B 1985. Chagas disease in pre-Colombian South America. *Amer J Physical Anthropology 68*: 495-498.

Sibley CG, Ahlquist JE 1984. The phylogeny of the hominoid primates, as indicated by DNA-DNA hybridization. *J Mol Evol 20*: 22-15.

Smith AG, Smith DG, Funnell BM 1994. *Atlas of Mesozoic and Cenozoic Coastlines*, Cambridge University, Cambridge, 99 pp.

Sogin ML, Elwood HJ, Gunderson JH 1986. Evolutionary diversity of eukaryotic small-subunit rRNA genes. *Proc Natl Acad Sci USA 83*: 1383-1387.

Stevens JR, Gibson WC 1998. The evolution of Salivarian trypanosomes. Second World Meeting on Salivarian Trypanosomes, Tryplink internet conference, EMBRAPA/ Centro de Pesquisa Agropecuaria do Pantanal, Brazil.

Stevens JR, Noyes HA, Dover GA, Gibson WC 1998. The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. *Parasitology* (in press).

Swofford DL, Olsen GJ, Waddell PJ, Hillis DM 1996. Phylogenetic inference, p. 407-514. In DM Hillis, C Moritz, BK Mable (eds), *Molecular Systematics*, Sinauer Associates, Sunderland, MA.

Vickerman K 1994. The evolutionary expansion of the trypanosomatid flagellates. *Int J Parasitol 24*: 1317-1331.

Weinman D 1972. *Trypanosoma cyclops* n. sp.: a pigmented trypanosome from the Malaysian primates *Macaca nemestrina* and *M. ira*. *Trans R Soc Trop Med Hyg 66*: 628-636.

Wiemer EAC, Hannaert V, Van Den Ijssel PRLA, Van Roy J, Opperdoes FR, Michels PAM 1995. Molecular analysis of glyceraldehyde-3-phosphate dehydrogenase in *Trypanosoma borelli*: An evolutionary scenario of subcellular compartmentation in Kinetoplastida. *J Mol Evol 40*: 443-454.

Wilson AC, Ochman H, Prager EM 1987. Molecular time scale for evolution. *Trends Genet 3*: 241-247.

Zuckerkandl E, Pauling L 1965. Evolutionary divergence and convergence in proteins, p. 97-166. In V Bryson, HJ Vogel (eds), *Evolving Genes and Proteins*, Academic Press, New York.

Genetic Data Showing Evolutionary Links between *Leishmania* and *Endotrypanum*

Elisa Cupolillo/+, Luiza OR Pereira, Octávio Fernandes*, Marcos P Catanho*, Júlio C Pereira, Enrique Medina-Acosta**, Gabriel Grimaldi Jr**

Laboratório de Leishmaniose, Departamento de Imunologia *Departamento de Medicina Tropical, Instituto Oswaldo Cruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil **Laboratório de Biotecnologia, Universidade Estadual do Norte Fluminense, 28015-620 Campos, RJ, Brasil

Striking similarities at the morphological, molecular and biological levels exist between many trypanosomatids isolated from sylvatic insects and/or vertebrate reservoir hosts that make the identification of medically important parasites demanding. Some molecular data have pointed to the relationship between some Leishmania *species and* Endotrypanum, *which has an important epidemiological significance and can be helpful to understand the evolution of those parasites. In this study, we have demonstrated a close genetic relationship between* Endotrypanum *and two new leishmanial species,* L. (V.) colombiensis *and* L. (V.) equatorensis. *We have used (a) numerical zymotaxonomy and (b) the variability of the internal transcribed spacers of the rRNA genes to examine relationships in this group. The evolutionary trees obtained revealed high genetic similarity between* L. (V.) colombiensis, L. (V.) equatorensis *and* Endotrypanum, *forming a tight cluster of parasites. Based on further results of (c) minicircle kDNA heterogeneity analysis and (d) measurement of the sialidase activity these parasites were also grouped together.*

Key words: *Leishmania colombiensis* - *Leishmania equatorensis* - *Endotrypanum* - multilocus enzyme electrophoresis - molecular characterization - numerical analysis - sialidase activity - kDNA

Parasitic protozoa of the genus *Leishmania* (Kinetoplastida: Trypanosomatidae) are biologically diverse group of microorganisms. Taxonomic studies of leishmanial isolates from the New World indicate tremendous diversity within this genus (Cupolillo et al. 1995). A number of new *Leishmania* species from sylvan areas of the Neotropics are associated with disease in humans; others appear to be restricted to lower orders of mammals, such as rodents and edentates (Grimaldi et al. 1989).

Sloths are reservoir hosts of at least five named *Leishmania* species of the subgenus *Viannia* [*L. guyanensis* Floch, 1954; *L. panamensis* Lainson & Shaw 1972; *L. shawi* Lainson et al. 1989; *L. colombiensis* Kreutzer et al. 1991 and *L. equatorensis* Grimaldi et al. 1992], responsible for human cutaneous and/or mucosal leishmaniasis (Grimaldi & Tesh 1993). Infections with other biologically distinct groups of trypanosomatid protozoa, such as *Endotrypanum* and *Trypanosoma* are also found in sloths (Deane 1961, Pipkin 1968, Travi et al. 1989, Shaw 1992).

In nature, all *Leishmania* spp. are transmitted by the bite of infected phlebotomine sand flies (Diptera:Psychodidae). However, many flagellates other than *Leishmania* commonly are found in sand flies in Neotropical forests. Arias et al. (1985) identified *E. schaudinni* and other *Endotrypanum* sp. infections in sand flies and sloths captured in the Amazon Region of Brazil. Results of kinetoplast DNA probe identifications of promastigotes present in sand flies captured near Manaus, Brazil also demonstrated *Endotrypanum* infections in *Lu. shannoni*, as well as in *Lu. umbratilis* and *Lu. anduzei* (Rogers et al. 1988). Further evidence for the development of *Endotrypanum* in phlebotomines was obtained by feeding several laboratory-reared sand fly species on infected sloths (Christensen and Herrer 1976, 1979, Shaw 1981).

Endotrypanum spp. are digenetic trypanosomatids in that they are intraerythrocytic parasites of sloths and are transmitted by phlebotomine sand flies (Shaw 1992). *Endotrypanum* shares many other charecteristics with *Leishmania*. Cul-

This work was supported by grants of the Instituto Oswaldo Cruz-Fiocruz, FAPERJ and CNPq to EC, OF and GG Jr. Research work at the Laboratório de Biotecnologia-UENF is supported by grants from FENORTE, FINEP, FAPERJ and WHO/TDR to EM-A.
+Corresponding author. Fax: +55-21-280.1589. E-mail: ecupoli@gene.dbbm.fiocruz.br
Received 15 June 1998
Accepted 30 July 1998

tured-derived promastigotes of parasites in both genera are morphologically similar. Studies employing monoclonal antibodies for the analysis of the genus *Endotrypanum* have shown antigenic similarites between these parasites and some *Leishmania* species (Franco et al. 1997). Furthemore, molecular trees clustered the sandfly-borne digenetic parasites *Leishmania* and *Endotrypanum* together, sharing a common ancestor and representing a relatively recent lineage from the Trypanosomatidae family (Fernandes et al. 1993). Results of hybridization using kDNA probes (Pacheco et al. 1990) support the view that *Endotrypanum* and the peripylarian leishmanial parasites of the subgenus *Viannia* Lainson & Shaw 1987 are phylogenetically close (Shaw 1992). In addition, phylogenetic studies have demonstrated that the most divergent *Leishmania* species are *L. (L.) hertigi* and *L. (L.) herreri*, claimed to be closer to *Endotrypanum* than to the other *Leishmania* (Croan & Ellis 1996, Noyes et al. 1996, 1997, Croan et al. 1997).

In this study, we have shown evolutionary links between *Endotrypanum* and some leishmanial parasites based on their molecular genetics, as characterized using a broad assemblage of methodologies. The data presented here demonstrate that *E. schaudinni*, *L. (V.) colombiensis* and *L. (V.). equatorensis* form a tight phylogenetic cluster, an evolutionary linked group that should be explored to understand the origin(s) of neotropical pathogenic *Leishmania*.

MATERIALS AND METHODS

Parasites - *Leishmania* and *Endotrypanum* (Table I) were cultured in Scheneider's Drosophila medium (Gibco, Grand Island, NY) supplemented with 10% heat-inactivated FBS (Biolab, Rio de Janeiro, Brazil) at 24°C. In the preparation of samples, the parasite (promastigotes in the late phase of growth cultures) were harvested by centrifugation (3,800x g for 15 min at 4°C) and washed twice in saline pH 8.0, containing the appropriate buffer.

Biochemical/Molecular characterization - The procedures used for characterizing the parasites (multilocus enzyme electrophoresis - MLEE, measurement of the sialidase activity, PCR amplification and restriction enzyme digestion of the parasite ITSrRNA, cloning and sequencing of the conserved region of the minicircle kDNA molecules) have been described in detail in previous publications (Cupolillo et al. 1994, 1995, Medina-Acosta et al. 1994, Fernandes et al. 1996). Sialidase activity was measured using a single-cell HITACHI F-4500 spectrofluorometer (350 nm excitation and 460 nm emission wavelengths). The sequencing was performed in automatic sequencing (AbiPrisma, Applied Biosystem).

Numerical analysis - The MLEE data was analyzed by phenetic methods using the NTSYS software program (version 1.7, exeter software). Principal coordinate analysis was performed based on Euclidian distance between the samples. The similarity level between the *Leishmania* species and *Endotrypanum* was calculated using the Jaccard's coefficient. The kDNA sequences of the parasites were analyzed using the MEGA program (Kumar et al. 1993). The number of differences between the sequences were calculated and a similarity tree constructed by the Neighbor-Joining method. Bootstrap analysis was based on 500 replicates.

RESULTS AND DISCUSSION

Leishmania and *Endotrypanum* are very close protozoan parasites (Fernandes et al. 1993) commonly found in the same vertebrate and insect hosts. Recents studies have showing the relationship between *Endotrypanum* and some New World *Leishmania* species, mainly those from *L. (L.) hertigi* and *L. (L.) herreri* complex (Croan & Ellis 1996, Noyes et al. 1996, 1997, Croan et al. 1997). Moreover, DNA analysis of phylogenetically informative RNA polymerase II gene of *L. (V.) equatorensis* and *Endotrypanum* demonstrated sequence similarities among these parasites (JJ Shaw, pers. commun.). Similarly, it appears that a close antigenic links may exist between *L. (V.) colombiensis*, *L. (V.) equatorensis* and *Endotrypanum* (Franco et al. 1997, Grimaldi et al. 1992).

Leishmania (V.) colombiensis was found infecting humans, sloths (*Choloepus hoffmanni*), sandflies (*Lu. hartmani* and *Lu. gomezi*), and dogs in Colombia, Panama, and Venezuela (Kreutzer et al. 1991, Delgado et al. 1993, unpublished data). *L. (V.) equatorensis* is an enigmatic parasite, which was isolated from the viscera of a sloth (*C. hoffmanni*) and a squirrel (*Sciurus granatensis*), captured in humid tropical forest on the Pacific Coast of Ecuador. Data based on biological and molecular criteria, as well as numerical zymotaxonomy analysis indicated that both these parasites are clearly distinguishable from all other known species, but clustered within the *L. (V.) braziliensis* complex (Kreutzer et al. 1991, Grimaldi et al. 1992). Multilocus enzyme electrophoresis data and the restriction fragments of the internal transcribed spacers of the rRNA gene (Cupolillo et al. 1995, 1997) have indicated a close relationship between *L. (V.) equatorensis* and *L. (V.) colombiensis*, as previously demonstrated (Kreutzer et al. 1991, Grimaldi et al. 1992). In order to better understand their taxonomic position in the genus, especially in relation to the discrimi-

TABLE I

Origin and Identification of *Leishmania* and *Endotrypanum* strains used in this study

Stock number	Designation[a]	Species	Geographic origin
L565	MHOM/BR/75/M4147	*L. guyanensis*	Brazil, Pará
L566	MHOM/BR/00/M2903	*L. braziliensis*	Brazil, Pará
L575	IFLA/BR/67/PH8	*L. amazonensis*	Brazil, Pará
L579	MHOM/BR/74/PP75	*L. chagasi*	Brazil, Bahia
L888	MCHO/EC/82/Lsp1[a]	*L. equatorensis*	Ecuador, Guayas
L889	MSCI/EC/82/Lsp2	*L. equatorensis*	Ecuador, Guayas
L1023	MHOM/BR/81/M6426	*L. lainsoni*	Brazil, Pará
L1245	IGOM/PA/85/E582.34	*L. colombiensis*	Panama, Colon
L1246	IPAN/PA/85/E696.26	*L. colombiensis*	Panama, Colon
L1247	IGOM/PA/85/E582.36	*L. colombiensis*	Panama, Colon
L1545	MHOM/BR/84/M8408	*L. shawi*	Brazil, Pará
L1365	MDAS/BR/79/M5533	*L. naiffi*	Brazil, Pará
E14	MCHO/BR/80/M6159[b]	*E. schaudinni*	Brazil, Pará

a: host [M=Mammalia: CHO=*Choloepus* sp. ([a]*C. hoffmanni*, [b]*C. didactylus*), DAS=*Dasypus novemcinctus*, HOM=*Homo sapiens*, SCI=*Sciurus granatensis*; I=Insecta: FLA=*Lutzomyia flaviscultelata*, GOM= *Lu. gomezi*, PAN=*Lu. panamensis*]/country of origin/year of isolation/original code.

nation of *Leishmania* from *Endotrypanum* and evolutive studies we decided to analyze the genetic similarity among these parasites, using several biochemical and molecular methods. This information will help define the fundamental mechanisms involved in species identification and taxonomic divergence among these microorganism.

The sialidase (EC 3.2.1.18) activity alone has been shown to be a good marker to discriminate between morphologically indistinguishable flagellates isolated from human, insects and sylvatic vertebrate reservoir hosts, such as *Leishmania* and *Endotrypanum* (Medina-Acosta et al. 1994). The general concensus is that *Endotrypanum* reference stocks express clear-cut varying levels of sialidase activities whereas the *Leishmania* reference stocks do not. In this study, we measured the sialidase activity for several neotropical *Leishmania* species and for reference strain *E. schaudinni.* As expected, *Endotrypanum* exhibited high levels of sialidase activity, whilst the taxonomically unquestionable *Leishmania* stocks (*i.e., L. chagasi*) were negative for this activity. However, high levels of sialidase activity were consistently obtained from both cell lysates and culture supernatants of *L. (V.) colombiensis* and *L. (V.) equatorensis*, levels comparable with those obtained for *E. schaudinni* (this work) and those of *Trypanosoma rangeli* and *Trypanosoma leeuwenhoeki* (Medina-Acosta et al. 1994).

Further, MLEE analyses demonstrated that *L. (V.) colombiensis* and *L. (V.) equatorensis* share alleles with *Endotrypanum* for some loci, such as G6PDH and IDHNAD, that were previously admitted as monomorphic for the latter genus and as discriminative characters between *Leishmania* and *Endotrypanum* (Franco et al. 1996). Moreover, for the malic enzyme were found two distinct loci (ME1 and ME2) for *L. (V.) equatorensis* and *L. (V.) colombiensis*, as described for *Endotrypanum* but in contrast to other leishmanial parasites (Cupolillo et al. 1994, Franco et al. 1996). According to the phenetic analyses, the results showed a high level of similarity between the two *Leishmania* species, as well as a close relationship between this group and *Endotrypanum* (Fig. 1, Table II). The later parasite is genetically closest to *L. (V.) colombiensis* rather than to *L. (V.) equatorensis* (Table II). In addition, the clusters *L. braziliensis/L. naiffi* and *L. guyanensis/L. shawi* were observed, as already demonstrated (Cupolillo et al. 1994, 1997) and *L. lainsoni* made a link between *L. (V.) colombiensis/L. (V.) equatorensis/E. schaudinni* and the *Leishmania (Viannia)* species.

The Neighbor-Joining tree constructed based on kDNA sequence data using the number of differences between *Leishmania* and *Endotrypanum* shows similar clustering of MLEE for *L. (V.) equatorensis/L. (V.) colombiensis/E. schaudinni* (Fig. 2). The position of *L. lainsoni* was maintained, forming a link between the group *L. (V.) equatorensis/L. (V.) colombiensis/E. schaudinni* and other *Leishmania* species. *Leishmania (V.) lainsoni* represents a very divergent monophyletic *Viannia* species, which was clustered as an independent complex (Thomaz-Soccol et al. 1993, Cupolillo et al. 1994, Fernandes et al. 1995, Eresh et al. 1995).

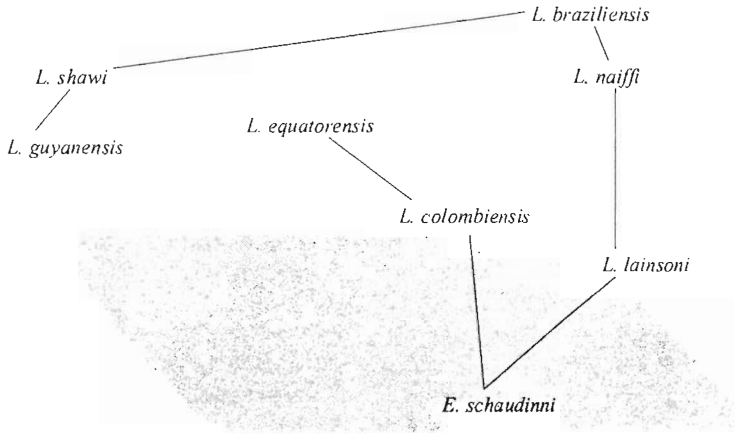


Fig. 1: principal coordinate analysis of the multilocus enzyme electrophoresis data. The three principal coordinates were claculated by Euclidian distance and plotted in 3D scale (the three principal coordinates represent 67.51% of the total variance). A minimum spanning tree was superimposed on the ordinations.

TABLE II

Similarity level among *Leishmania* and *Endotrypanum* species calculated by the Jaccard's coefficient

	1	2	3	4	5	6	7	8
1. *L. braziliensis*	-							
2. *L. guyanensis*	0.21	-						
3. *L. lainsoni*	0.22	0.16	-					
4. *L. equatorensis*	0.13	0.08	0.11	-				
5. *L. colombiensis*	0.16	0.07	0.14	0.40	-			
6. *E. schaudinni*	0.10	0.13	0.08	0.18	0.23	-		
7. *L. shawi*	0.24	0.59	0.13	0.07	0.05	0.07	-	
8. *L. naiffi*	0.47	0.15	0.26	0.12	0.18	0.12	0.20	-

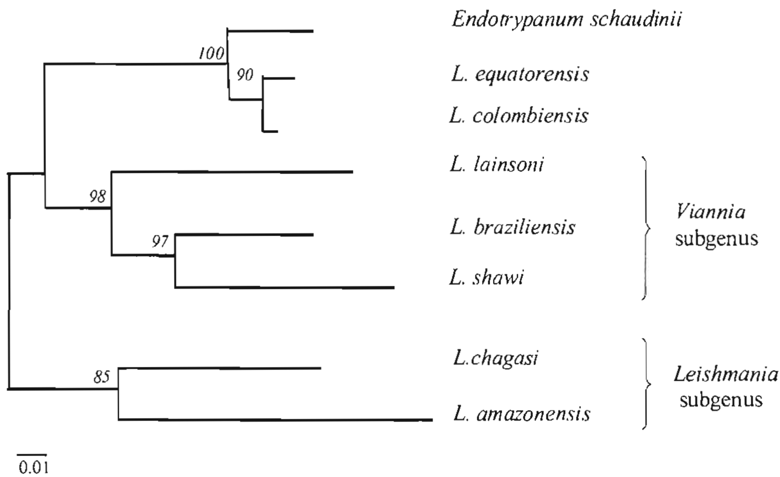


Fig. 2: phenetic analyze of sequences (83bp) of conserved region of kDNA minicircle. The similarities were evaluated by the number of differences among the sequences and the similarity tree constructed by the Neighbor-joining method. Italic numbers represent bootstrap values based on 500 replicates.

The internal transcribed spacers of the rRNA gene were amplified by PCR and the product digested with several restriction enzymes (Cupolillo et al. 1995). The RFLP profiles show a close but not identical pattern between *L. (V.) colombiensis* and *L. (V.) equatorensis*. However, through this method *Endotrypanum* can be easily discriminated from the former parasites and other *Leishmania* species by most of the restriction enzyme profiles (Fig. 3).

The genetic similarity between *Endotrypanum* and New World *Leishmania* was also demonstrated by sequencing comparisons of the small subunit of ribosomal RNA and RNA Polymerase II genes (Croan & Ellis 1996, Noyes et al. 1996, 1997, Croan et al. 1997). The results show that *L. (L.) herreri* (Zeledon et al. 1975), a sloth parasite, is closer to *Endotrypanum* than to other *Leishmania* species. *Leishmania (L.) hertigi/L. (L.) deanei* (Herrer 1971, Lainson & Shaw 1977), which were isolates from rodents, are also genetically closest to the *Endotrypanum/L. (L.) herreri* group (Croan et al. 1997, Noyes et al. 1997). Some authors suggest that *L. (L.) herreri* is a misclassified parasite and therefore probably represents *Endotrypanum* (Croan & Ellis 1996). Although *L. (L.) hertigi* and *L. (L.) deanei* are still enigmatic parasites (Lainson 1997) there are evidences supporting their classification as *Leishmania*. An interesting aspect is that these *Leishmania* species and *Endotrypanum* are biologically distinct parasites and do not share the same hosts.

In contrast to *L. (V.) colombiensis*, which has been isolated from humans (Kreutzer et al. 1991, Delgado et al. 1993), the public health importance of *L. (V.) equatorensis* remains to be determined. To date, it has only been isolated from arboreal mammals; no human infections with the parasite have been identified. Likewise, the sandfly vector (s) are unknown. However, the biological behaviour of *L. (V.) equatorensis* is indistinguishable from other members of the *L. (V.) braziliensis* complex, based on its virulence and development in laboratory animals. Inoculation of cultured promastigotes into the nose of hamster (*Mesocricetus auratus*) produced local swelling without metastasis; appearance of the lesions took 1-3 months, depending on the size of the inoculum (Grimaldi et al. 1992). Moreover, the restriction profile of the internal transcribed spacers of the rRNA gene showed a close pattern between *L. (V.) equatorensis* and *L. (V.) colombiensis,* but distinct from *Endotrypanum*, supporting the taxonomic status of the former parasite, and that the two *Leishmania* species represent a link between *Endotrypanum* and *Leishmania*.

Comparative studies will be needed to address the antiquity of this evolutionary link group and, in particular, whether or not it represents a branch point on the origin of neotropical leishmanias. It is worth noting that sloths, which have always been restricted to the American continent, are consid-

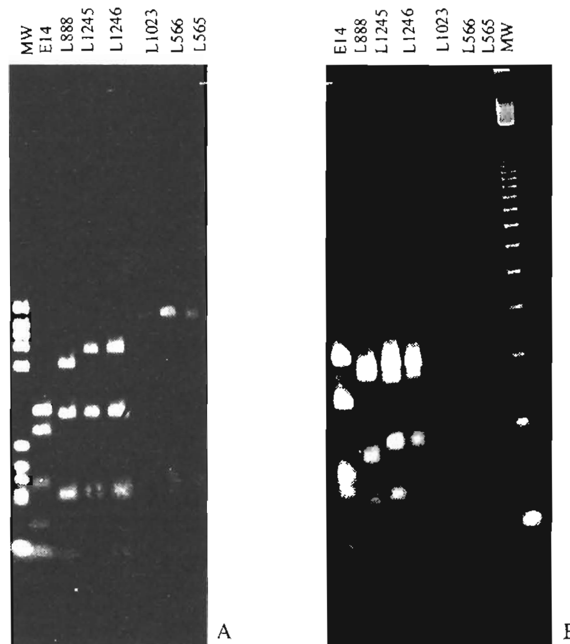


Fig. 3: restriction enzyme profile of the internal transcribed spacers of the rRNA genes for *Leishmania* species and *Endotrypanum schaudinni*. A. *Bst*UI; B. *Taq* I.

ered to have evolved from the basic Xenarthran armadillo-like stock some 60 million years ago during the Palaeocene period. These early mammals separated between the two and the three-toed groups of extant sloths later during the Miocene period. With this in mind, we feel that the neotropical leishmanias may well have evolved from a primitive endotrypanumal Miocene parasite line of South American sloths.

REFERENCES

Arias JR, Miles MA, Naiff RD, Póvoa MM, de Freitas RA, Biancardini CB, Castellon EG 1985. Flagellate infections of Brazilian sand flies (Diptera: Psychodidae): isolation *in vitro* and biochemical identification of *Endotrypanum* and *Leishmania*. *Am J Trop Med Hyg 34:* 1096-1108.

Christensen H, Herrer A 1976. Neotropical sandflies (Diptera:Psychodidae), invertebrate hosts of *Endotrypanum schaudinni* (Kinetoplastida). *J Med Entomol 13*: 299-303.

Christensen H, Herrer A 1979. Susceptability of sandflies (Diptera:Psychodidae) to trypanosomatidae from two-toed sloths (Edentata: Bradypodidae). *J Med Entomol 16*: 424-427.

Croan D, Ellis J 1996. Phylogenetic relationships between *Leishmania*, *Viannia* and *Sauroleishmania* inferred from comparison of a variable domain within the RNA polymerase I I largest subunit gene. *Mol Biochem Parasitol 79*: 97-102.

Croan DG, Morrison DA, Ellis JT 1997. Evolution of the genus *Leishmania* revealed by comparison of DNA and RNA polymerase gene sequences. *Mol Biochem Parasitol 89*: 149-159.

Cupolillo E, Grimaldi Jr G, Momen H 1994. A general classification of New World *Leishmania* using numerical zymotaxonomy. *Am J Trop Med Hyg 50:* 296-311.

Cupolillo E, Grimaldi Jr G, Momen H 1997. Genetic diversity among *Leishmania (Viannia)* parasites. *Ann Trop Med Parasitol 91*: 617-626.

Cupolillo E, Grimaldi Jr G, Momen H, Beverley SM 1995. Intergenic region typing (IRT): a rapid molecular approach to the characterization and evolution of *Leishmania. Mol Biochem Parasitol 73:* 145-155.

Deane LM 1961. Tripanosomatídeos de mamíferos da região Amazônica. I. Alguns flagelados encontrados no sangue de mamíferos silvestres do Estado do Pará. *Rev Inst Med Trop São Paulo 3:* 15-28.

Delgado O, Castes M, White Jr AC, Kreutzer RD 1993. *Leishmania colombiensis* in Venezuela. *Am J Trop Med Hyg 48:* 145-147.

Eresh S, de Brujin MHL, Mendoza-León JA, Barker DC 1995. *Leishmania (Viannia) lainsoni* occupies a unique niche within the subgenus *Viannia*. *Trans R Soc Trop Med Hyg 89:* 231-236.

Fernandes AP, Nelson K, Beverley SM 1993. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: Perspectives on the age and origins of parasitism. *Proc Natl Acad Sci USA 90:* 11608-11612.

Fernandes O, Bozza M, Pascale JM, Miranda AB, Lopes UG, Degrave WM 1996. An oligonucleotide probe derived from kDNA minirepeats is specific for *Leishmania (Viannia)*. *Mem Inst Oswlado Cruz 91*: 279-284.

Fernandes O, Pacheco RS, Momen H, Degrave W, Campbell DA 1995. *Leishmania lainsoni:* a peculiar *Viannia* species. *Ann Trop Med Hyg 89:* 81-82.

Floch H 1954. *Leishmania tropica guyanensis* n.sp. agent de la leishmaniose tegumentarie de Guyanes et de l'Amerique Centrale. *Arch Inst Pasteur de La Guyane Française et du Teritoire de L'Inni 15:* 328.

Franco AMR, Machado GMC, Naiff RD, Moreira CFS, McMahon-Pratt D, Grimaldi Jr G 1997. Characterization of *Endotrypanum* parasites using specific monoclonal antibodies. *Mem Inst Oswaldo Cruz 92*: 63-68.

Franco AMR, Momen H, Naiffi RD, Moreira CFS, Deane MP, Grimaldi Jr G 1996. Enzyme polymorphism in *Endotrypanum* and numerical analysis of isoenzyme data. *Parasitology 113*: 39-48.

Grimaldi Jr G, Tesh RB 1993. *Leishmania*ses of the New World: current concepts and implications for future research. *Clin Microbiol Rev 6:* 230-250.

Grimaldi Jr G, Kreutzer RD, Hashigushi Y, Gomez EA, Mimory T, Tesh RB 1992. Description of *Leishmania equatorensis* sp.n. (Kinetoplastida: Trypanosomatidae), a new parasite infecting arboreal mammals in Ecuador. *Mem Inst Oswaldo Cruz 87:* 221-228.

Grimaldi Jr G, Tesh RB, McMahon-Pratt D 1989. A review of the geographic distribution and epidemiology of *Leishmania*sis in the New World. *Am J Trop Med Hyg 41:* 687-725.

Herrer A 1971. *Leishmania hertigi* sp. n. from the tropical porcupine, *Coendou rothschildi* Thomas. *J Parasitol 57:* 626-629.

Kreutzer RD, Corredor A, Grimaldi Jr G, Grogl M, Rowton ED, Young DG, Morales A, McMahon-Pratt D, Guzman H, Tesh RB 1991. Characterization of *Leishmania colombiensis* sp.n. (Kinetoplastida: Trypanosomatidae), a new parasite infecting humans, animals, and phlebotomine sand flies in Colombia and Panama. *Am J Trop Med Hyg 44:* 662-675.

Kumar S, Tamura K, Nei M 1993. MEGA: Molecular evolutionary analysis, version 1.01. The Pennsylvania State University, Philadelphia, PA.

Lainson R 1997. *On Leishmania enriettii* and other enigmatic *Leishmania* species of the neotropics. *Mem Inst Oswaldo Cruz 92*: 377-387.

Lainson R, Shaw JJ 1972. *Leishmania*sis in the New World: taxonomic problems. *Brit Med Bull 28:* 44-48.

Lainson R, Shaw JJ 1977. *Leishmania* of neotropical porcupines: *Leishmania hertigi deanei* nov. subsp. *Acta Amazonica 7*: 51-57.

Lainson R, Braga RR, De Souza AAA, Póvoa MM, Ishikawa EAY, Silveira FT 1989. *Leishmania (Viannia) shawi* sp. n., a parasite of monkeys, sloths and procyonids in Amazonian Brazil. *Ann Parasitol Hum Comp 64*: 200-207.

Medina-Acosta E, Franco AMR, Jansen AM, Sampol

M, Nevés N, Pontes-de-Carvalho L, Grimaldi Jr G, Nussenzweig V 1994. Trans-sialidase and sialidase activities discriminate between morphologically indistinguishable trypanosomatids. *Eur J Biochem 225:* 333-339.

Noyes HA, Arana BA, Chance ML, Maingon R 1997. The *Leishmania hertigi* (Kinetoplastida; Trypanosomatidae) complex and the lizard *Leishmania*: their classification and evidence for a neotropical origin of the *Leishmania-Endotrypanum* clade. *J Euk Microbiol 44*: 511-517.

Noyes HA, Camps AP, Chance ML 1996. *Leishmania herreri* (Kinetoplastida; Trypanosomatidae) is more closely related to *Endotrypanum* (Kinetoplastida; Trypanosomatidae) than to *Leishmania*. *Mol Biochem Parasitol 80*: 119-123.

Pacheco RS, Thomaz N, Momen H 1990. kDNA cross-hybridization between *Endotrypanum* and peripylarian *Leishmania*. *Trans R Soc Trop Med Hyg 84*: 531.

Pipkin AC 1968. Domicillary reduviid bug and the epidemiology of Chagas'disease in Panama (Hemiptera: Reduviidae: Triatominae). *J Med Entomol 5:* 107-124.

Rogers WO, Burnheim PF, Wirth DF 1988. Detection of *Leishmania* within sandflies by kinetoplast DNA hybridization. *Am J Trop Med Hyg 39:* 434-439.

Shaw JJ 1981. The behaviour of *Endotrypanum schaudinni* (Kinetoplastida: Trypanosomatidae) in three species of laboratory-bred neotropical sandflies (Diptera:Psychodidae) and its influence of the classification of the genus *Leishmania*, p. 232-241. In EU Canning, *Parasitological Topics*, Allen, Lawrence, KS.

Shaw JJ 1992. *Endotrypanum*, a unique intraerythrocytic flagellate of New World tree sloths. An evolutionary link or an evolutionary backwater? *Ciên Cult 44:* 107-116.

Thomaz-Soccol V, Lanotte G, Rioux J-A, Pratlong F, Martini-Dumas A, Serres E 1993. Monophyletic origin of the genus *Leishmania* Ross, 1903. *Ann Parasitol Hum Comp 68:* 107-108.

Travi BL, Zea A, D'Alessandro A 1989. *Trypanosoma (Herpetosoma) leeuwenhoeki* in *Choloepus hoffmanni* and *Didelphis marsupialis* of the pacific coast of Colombia. *J Parasitol 75:* 218-224.

Zeledon R, Ponce C, De Ponce E 1975. The isolation of *Leishmania braziliensis* from sloths in Costa Rica. *Am J Trop Hyg 24:* 706-707.

A Study of *Cryptosporidium parvum* Genotypes and Population Structure

G Widmer/+, L Tchack, F Spano*, S Tzipori

Tufts University School of Veterinary Medicine, Division of Infectious Diseases, North Grafton, Massachusetts, USA *Istituto di Parassitologia, Università di Roma "La Sapienza", Rome, Italy

Genetic evidence for the occurrence of two Cryptosporidium parvum *subgroups is presented. This evidence is based on restriction fragment length polymorphism analysis of several independent loci. Sequence analysis of the β -tubulin intron revealed additional polymorphism. The stability of the genetic profiles following passage of* C. parvum *isolates between different hosts was investigated.*

Key words: *Cryptosporidium parvum* - restriction fragment length polymorphism - tubulin

Cryptosporidium parvum is an enteric protozoan parasite which commonly infects immunosuppressed individuals. Ruminants, in particular calves, are important reservoirs. Recent genotypic analyses of *C. parvum* from human cases of cryptosporidiosis have identified two groups of genotypically distinct parasites. One of these genotypes, designated genotype C, infects animals and humans, whereas the other, known as genotype H, is only found in humans. Differences in infectivity between H and C isolates were found in animal models. These observations have led to the hypothesis that *C. parvum* is transmitted via different transmission routes, each transmitting parasites of one genotype. An alternative view is that both genotypes circulate among different host species, and that genotypically different populations can arise from mixed infections through selection in different host environments.

MATERIALS AND METHODS

DNA purification - PCR amplification was performed either on DNA isolated directly from stool or extracted from purified oocysts. For stool DNA extraction, 100 to 200 µl of stool was incubated overnight in 0.2% SDS and 200 µg/ml proteinase K, extracted with phenol/chloroform and ethanol precipitation. Alternatively, oocysts were purified from stool and DNA recovered by proteinaseK/SDS treatment.

Restriction fragment length polymorphism - Multilocus RFLP was performed using four unlinked RFLP markers; polyT (Carraway et al. 1997), COWP (Spano et al. 1997), TRAP-C1 (Spano et al. 1998) and RNR (Widmer et al. 1998). A sequence-specific PCR assay aimed at the ribosomal internal transcribed spacer 1 (Carraway et al. 1996) was also used.

Isopycnic fractionation of oocysts- Semi-purified oocysts were sedimented on a 15-30% (w/v) Nycodenz (Sigma) for 1 hr at 55,000xg. Fractions of approximately 1 ml were recovered and oocysts concentrated by centrifugation.

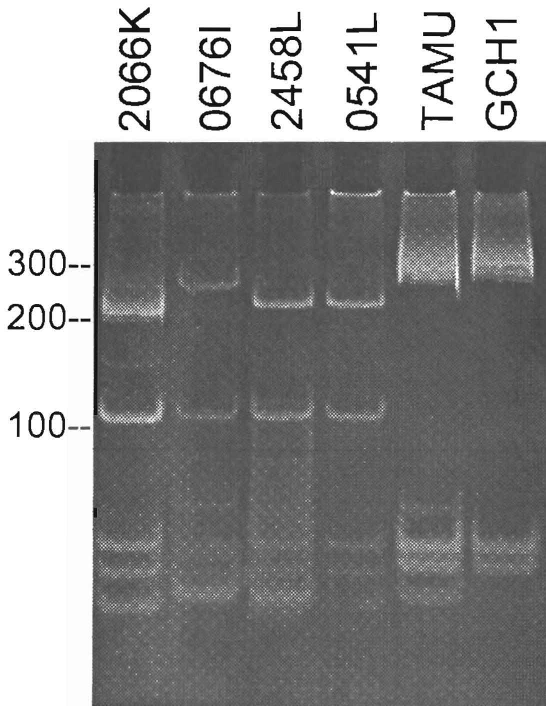
RESULTS AND DISCUSSION

In order to investigate the epidemiology of *C. parvum*, we have developed PCR and PCR-RFLP markers. Several coding and a non-coding region were examined for sequence polymorphism. Using a combination of polymorphic markers developed in our laboratories, *C. parvum* isolates originating from different host species and different geographical locations were subject to a multilocus genotypic analysis. Isolates were found to segregate into H (41%), C (52%) and mixed (7%) type isolates. Significantly, in a sample of 29 isolates no recombinants were identified, suggesting reproductive isolation between H and C parasites.

RFLP and sequence analysis of a non-coding region (the β-tubulin intron) identified a high degree of polymorphism (Fig.). A multiple sequence alignment of cloned PCR products spanning the β-tubulin intron and part of exon 2, revealed four groups of sequences and additional polymorphism within groups. Sequences indicative of interallelic recombination were found in two isolates.

The population structure of isolates serially transmitted through calves or passaged from calves to mice, human to mice or calves to humans was examined. Several infections showing changes in RFLP profiles following serial transmission were observed. Using isopycnic fractionation of oocysts, it was possible to separate, in the calf-propagated isolate GCH1, two populations of oocysts bearing different genotypes.

+Corresponding author. Fax: +508-839.7977. E-mail: gwidmer@infonet.tufts.edu
Received 15 June 1998
Accepted 30 July 1998



Restriction site polymorphism in the beta-tubulin gene of *Cryptosporidium parvum.* PCR products amplified from the intron and adjacent exon 2 were digested with restriction enzyme Tsp5091. Three restriction profiles were detected among these samples; one in the bovine isolate GCH1 and TAMU, one in human isolate 06761 and one in human isolates 2066K, 2458L and 0541L.

These observations indicate that the epidemiology of *C. parvum* is complex and that individual hosts can excrete heterogeneous populations of oocysts. The significance of these findings for human cryptosporidiosis has not been elucidated. Of primary interest is the question whether isolates of genotype H and C differ in virulence and susceptibility to drug treatment.

REFERENCES

Carraway M, Tzipori S, Widmer G 1996. Identification of genetic heterogeneity in the *Cryptosporidium parvum* ribosomal repeat. *Appl Environ Microbiol 62*: 712-716.

Carraway M, Tzipori S, Widmer G 1997. New RFLP marker in *Cryptosporidium parvum* identifies mixed parasite populations and genotypic instability in response to host change. *Infect Immun 65*: 3958-3960.

Spano F, Putignani L, McLaughlin J, Casemore DP, Crisanti A 1997. PCR-RFLP analysis of the *Cryptosporidium* oocyst wall protein (COWP) gene discriminates between *C. wrairi* and *C. parvum*, and between *C. parvum* isolates of human and animal origin. *FEMS Microbiol Letters 152*: 209-217.

Spano F, Putignani L, Naitza S, Puri C, Wright S, Crisanti A 1998. Molecular cloning and expression analysis of a *Cryptosporidium parvum* gene encoding a new member of the thrombosponding family. *Molec Biochem Parasitol 92*: 147-162.

Widmer G, Tzipori S, Fichtenbaum CJ, Griffiths JK 1998. Genotypic and phenotypic characterization of *Cryptosporidium parvum* isolates from people with AIDS. *J Infect Dis:* in press.

Species and Strain-specific Typing of *Cryptosporidium* Parasites in Clinical and Environmental Samples

Lihua Xiao/+, Irshad Sulaiman, Ronald Fayer*, Altaf A Lal

Division of Parasitic Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, US Department of Health and Human Services, Atlanta, GA 30341 *Parasite Immunobiology Laboratory, Agriculture Research Service, U.S. Department of Agriculture, Beltsville, MD 20705, USA

Cryptosporidiosis has recently attracted attention as an emerging waterborne and foodborne disease as well as an opportunistic infection in HIV infected individuals. The lack of genetic information, however, has resulted in confusion in the taxonomy of Cryptosporidium *parasites and in the development of molecular tools for the identification and typing of oocysts in environmental samples. Phylogenetic analysis of the small subunit ribosomal RNA (SSU rRNA) gene has shown that the genus* Cryptosporidium *comprises several distinct species. Our data show the presence of at least four species:* C. parvum, C. muris, C. baileyi *and* C. serpentis (C. meleagridis, C. nasorum *and* C. felis *were not studied). Within each species, there is some sequence variation. Thus, various genotypes (genotype 1, genotype 2, guinea pig genotype, monkey genotype and koala genotype, etc.) of* C. parvum *differ from each other in six regions of the SSU rRNA gene. Information on polymorphism in* Cryptosporidium *parasites has been used in the development of species and strain-specific diagnostic tools. Use of these tools in the characterization of oocysts in various samples indicates that* C. parvum *genotype 1 is the strain responsible for most human* Cryptosporidium *infections. In contrast, genotype 2 is probably one of the major sources for environmental contamination, and has been found in most oysters examined from Chesapeake Bay that may serve as biologic monitors of estuarine waters.*

Key words: *Cryptosporidium* - phylogeny - genotype - ribosomal RNA

Cryptosporidiosis is a coccidian infection of humans, domestic animals and other vertebrates. In young farm animals, especially preweaned dairy calves, it causes a severe enteritis resulting in significant morbidity, mortality and economic loss. In humans, it results in an acute infection of the digestive system in immunocompetent individuals, and chronic, life-threatening disease in immunocompromised patients. Several transmission routes, including person-to-person, contamination of water or food, and zoonotic infection, are possible. The specific source of *Cryptosporidium* oocysts involved in infection or contamination is frequently unknown, largely due to a lack of detailed epidemiologic investigation and strain-typing tools. The latter results from a current paucity of molecular characterization and lack of acceptance of the taxonomy of *Cryptosporidium* species and genotypes.

CRYPTOSPORIDIUM SPECIES

Since the discovery of *Cryptosporidium muris* and *C. parvum* in rodents, over 20 *Cryptosporidium* species have been described in various animal hosts (O'Donoghue 1995). Species were named based on the historical belief that *Cryptosporidium* spp. are coccidian parasites, and therefore share the strict host specificity demonstarted by many other coccidian parasites. Studies conducted in late 1970s and early 1980s, however, indicated that some isolates of *Cryptosporidium* were infectious for several animal species. Thus, one group of investigators suggested that all *Cryptosoridium* parasites were the same species, *C. muris* (Tzipori et al. 1980). Others demonstrated that host specificity was present among isolates from different classes of vertebrates (O'Donoghue 1995). Based on these observations, Levine (1984, 1986) classified the parasites from mammals, birds, reptiles and fish as *C. muris*, *C. meleagridis*, *C. serpentis*, and *C. nasorum*, respectively. Subsequent studies demonstrated that *C. parvum* from mammals and *C. baileyi* from birds were biologically and morphologically different from *C. muris* and *C. meleagridis*

This work was supported in part by inter-agency agreements (#DW75937730-01-0 and DW7593784-01-0) from CDC and EPA, and Emerging Infectious Diseases and Opportunistic Infectious Diseases funds from CDC, USA.
+Corresponding author. Fax: +770-488-4454.
Received 15 June 1998
Accepted 30 July 1998

(Upton & Current 1985, Current et al. 1986). Thus, *C. parvum*, *C. muris*, *C. baileyi*, *C. meleagridis*, *C. serpentis* and *C. nasorum* were considered valid *Cryptosporidium* species (O'Donoghue 1995). More recently, based on published reports of host specificity, Fayer et al. (1997) added *C. felis* from cats and *C. wrairi* from guinea pigs to the list of valid species, whereas Tzipori and Griffiths (1998)suggested that current evidence does not support the concept that there is more than one species of *Cryptosporidium* parasites.

The lack of genetic information and the presence of erroneous sequences in a few published studies have added to the present state of taxonomic confusion. Cai et al. (1992) compared the small subunit (SSU) ribosomal RNA (rRNA) gene, and showed a greater than 99% identity between one *C. parvum* and one *C. muris* isolate. Alignment of sequences (accession numbers X64430 to 64343) from that study with sequences from us and others indicates that all four sequences from Cai et al. (1992) are the *C. muris* type. Minor sequence errors (one insertion and 12 deletions of nucleotides) were found in the SSU rRNA sequence (L25642) of another published study (Kilani & Wenman 1994). These sequences and five other sequences deposited in the GenBank were used recently by Tzipori and Griffiths (1998) in a phylogenetic analysis of *Cryptosporidium* parasites. Based on this analysis, they concluded that the observed inter-species and intra-species variation did not favor the designation of separate *Cryptosporidium* species, and therefore all *Cryptosporidium* oocysts, including those from lower vertebrates, should be considered hazardous to humans.

We have recently sequenced the SSU rRNA genes from various isolates of *C. parvum*, *C. muris*, *C. baileyi* and *C. serpentis*, and used these sequences in a phylogenetic analysis (Xiao et al. unpub. data). Results of the analysis indicate that *Cryptosporidium* parasites are a multi-species complex containing at least four species: *C. parvum*, *C. baileyi*, *C. muris* and *C. serpentis* (*C. felis*, *C. nasorum* and *C. meleagridis* were not studied). The evolutionary distance between the *Cryptosporidium* guinea pig isolate and *C. parvum* is too small to warrant a separate species designation.

CRYPTOSPORIDIUM PARVUM GENOTYPES

Results of various studies indicate that there is variation within the species *C. parvum*. Two dimensional gel electrophoresis has revealed minor differences between human and bovine *C. parvum* isolates (Mead et al. 1990), which has been confirmed by immunoblot (Nichols et al. 1991, Nina et al. 1992), isozyme (Ogunkolade et al. 1993, Awad-El-Kariem et al. 1995), and restriction fragment length polymorphism (RFLP) analysis (Ortega et al. 1991). More recently, random amplified polymorphic DNA (RAPD) markers have revealed two distinct groups of human *C. parvum* isolates, one containing most human isolates and the other containing some human isolates and all animal isolates (Morgan et al. 1995), indicating the possibility of zoonotic infection. Similar results have been obtained by sequence data or PCR-RFLP analysis of a repetitive sequence (Bonnin et al. 1996), bifunctional dihydrofolate reductase thymidylate synthase (DHFR) (Vasquez et al 1996), rRNA repeats (Carraway et al. 1996), polythreonine motifs (Carraway et al. 1997), oocyst wall protein (COWP) gene (Spano et al. 1997), and thrombospondin anonymous protein-2 (TRAP-C2) gene (Peng et al. 1997 Sulaiman et al. unpub. data). It remains unclear, however, whether the same two genotypes are present in all these polymorphic loci. Results of our multi-locus analysis suggest that indeed the same genotypes are linked across all polymorphic genes (SSU rRNA, TRAP-C1, TRAP-C2, CP15, and ß-tubulin intron) examined (Xiao et al. unpub. data).

Our phylogenetic analyses of the SSU rRNA gene have revealed diversities in *C. parvum* not previously observed (Table I). Human *C. parvum* isolates differ from bovine isolates in four regions of the SSU rRNA gene. Likewise, the *Cryptosporidium* isolate from guinea pigs (*C. wrairi*) also differs from the bovine isolates in four regions, two of which are the same polymorphic regions between the human and bovine genotypes, thus representing a third genotype of *C. parvum*. Partial sequences obtained from a monkey by us and from a koala by Morgan et al.(1997) indicate the presence of two additional genotypes. The difference between the human and bovine genotypes in nucleotides 689-699 has also been observed recently by Morgan et al. (1997). We, however, have observed that some human isolates have the sequence TTTTTT instead of TTTTTTTTTTT. Based on a partial SSU rRNA gene sequence, another group also identified a new *C. parvum* genotype (Carraway et al. 1994, 1996). The new genotype sequence (ICP), however, is identical to the *C. muris* bovine isolate (Xiao et al. unpub. data).

CRYPTOSPORIDIUM GENOTYPES IN CLINICAL SAMPLES

Results of the molecular characterization have been used by us in the development of molecular diagnostic tools. A PCR-RFLP technique based on the polymorphism in the TRAP-C2 gene was developed and used in the analysis of human clinical samples from various outbreak and non-outbreak cases (Sulaiman et al., unpub. data). Results of our

TABLE I

Differences among genotypes of *Cryptosporidium parvum* in the SSU rRNA gene

	Location of mutations in the SSU rRNA gene[a]					
Genotype	129-135	179-184	262-267	639-656	689-699	795-800
1	TTTTACT	AAACTC	AATTAA	AAAATATTTGATGAATA	TTTTTTTTTTT or TTTTTT	TTTTTT
2	TTT-ACT	AAACTC	ATTAAA	AAAATATTTGATGAATA	TATATTTT	TTTCTT
wrairi	TTT-ACT	AGGCCC	ATAAAT	ATAATATTTGAA-AATA	TATATTTTT	TTTCTT
Monkey	unknown	unknown	AATTAA	AATATATTTTGATGAATA	TTTTTTTTT	TTTTTT
Koala[b]	unknown	unknown	unknown	ATTATACTTTTTAAGGTG	TATTTTTTT	unknown

a: nucleotide positions in the aligned sequences of all *Cryptosporidium* species. Actual positions in individual sequences may vary slightly due to the introduction of gaps in the aligned sequences (1757 bp); *b*: based on the sequence by Morgan et al. (1997).

studies and those by others (Table II) indicate that anthroponotic organisms account for the majority of the cases and person-to-person transmission is likely to be an important transmission route of cryptosporidiosis in non-outbreak cases. This is evident from the large number of genotype 1 parasites in sporadic cases and HIV patients (Sulaiman et al., unpub. data). This is in agreement with some recent observations by others (Table II). Even in outbreak cases, many cryptosporidiosis outbreaks are caused by anthroponotic (genotype 1) parasites (such as the waterborne outbreaks in Milwaukee in 1993, Nevada in 1994, and Florida in 1995; the Atlanta day care outbreak in 1995, and the Washington outbreak in 1997). It is possible that genotype 2 parasites largely cause human infection through contamination of water or food or direct contact with infected animals, especially in rural areas. Examples are the Maine apple cider outbreak in 1993, the British Columbia waterborne outbreak in 1996, and the Pennsylvania multi-family outbreak in 1997. The reason for the high percentage of genotype 2 in AIDS patients (6/13 patients) in France (Bonnin et al. 1996) is not clear. Taken together, there are two distinct populations of *C. parvum* parasites, one cycling only in humans and one cycling predominantly in animals. The latter can cause human infections.

CRYPTOSPORIDIUM PARASITES IN ENVIRONMENTAL SAMPLES

One difficulty facing the investigation of waterborne outbreaks of cryptosporidiosis is the lack of a sensitive, specific diagnostic tool. Most of the current PCR diagnostic and genotyping tools are designed for analysis of clinical samples. Because they cannot differentiate *Cryptosporidium* species and have low sensitivities, they have limitations in the analysis of water samples. Two PCR-RFLP techniques based on the SSU rRNA gene have claimed to differentiate *C. parvum* from other *Cryptosporidium* parasites (Awad-El-Kariem et al. 1994, Leng et al. 1996). One technique (Leng et al. 1996) used conserved sequences for primers and therefore amplify the SSU rRNA gene of all eukaryotic organisms. The other technique (Awad-El-Kariem et al. 1994) used erroneous sequence by Cai et al. (1992) as primers, reducing the efficiency of amplification and making interpretation of the data difficult. Nor have the present genotyping techniques been subjected to cross-species testing, making interpretation of results from environmental samples that could contain non-*parvum Cryptosporidium* virtually impossible.

Based on sequence information on the SSU rRNA gene, we have developed a PCR-RFLP technique for both species identification and genotyping of *Cryptosporidium* parasites. Because the technique employs nested PCR and targets the multi-copied rRNA gene, it has sufficient sensitivity for use in environmental samples. We have used this technique in the analysis of *Cryptosporidium* oocysts recovered from the gill washings and hemolymph of oysters (*Crassostrea virginica*) collected from the Chesapeake Bay. We are interested in oysters because they are filter feeders that concentrate and accumulate *Cryptosporidium* oocysts they have removed from surface waters. The use of oysters enables investigators to avoid the poor recovery rate often associated with filtering hundreds of liters of water to determine the presence or absence of *Cryptosporidium* oocysts. Before applying our technique *Cryptosporidium* oocysts were morphologically identified in oysters, but the species of most of the oocysts was unconfirmed (Fayer et al. 1998).

TABLE II

Prevalence of genotype 2 in human clinical samples reported in various studies

Location	Sample source	# of samples	genotypes 1/2	Technique used	Reference
England & Guinea Bissau	Sporadic cases	11	10/1	Isozyme	Awad-El-Kariem et al. 1995
Western & South Australia	Sporadic cases	14	12/2	RAPD	Morgan et al. 1995
USA	Sporadic cases	3	2/1	ITS1 and SSU rRNA repeat	Carraway et al. 1996
Northeast France	HIV+ patients	13	6/7	PCR-RFLP of repetitive DNA	Bonnin et al. 1996
UK	Sporadic cases	7	5/2	PCR-RFLP of oocyst wall protein	Spano et al. 1997
Western Australia	Sporadic cases	32	28/4	PCR of RAPD fragment	Morgan et al., 1997
USA & Canada	Outbreaks & sporadic cases	16	13/3	TRAP-C2 sequencing	Peng et al. 1997
USA, Canada, India & Guatemala	Outbreaks & sporadic cases	50	42/8	PCR-RFLP of TRAP-C2	Sulaiman et al. unpub. data

Preliminary analysis of 65 pooled oyster samples using the SSU rRNA-based PCR-RFLP technique has shown the presence of *Cryptosporidium* oocysts in 26 samples. Twenty four of these positive samples were *C. parvum*, and each of the others was *C. baileyi* and *C. serpentis*. The majority of *Cryptosporidium* oocysts were of genotype 2 (22 samples), indicating animals maybe the most likely the source of most *Cryptosporidium* oocyst contamination in the Chesapeake Bay. Even though this is a highly populated area, only two samples had genotype 1 sequences . These results demonstrate that oysters can serve as a biologic monitor for *Cryptosporidium* oocyst contamination in waters. Because raw oysters are often consumed by humans, *Cryptosporidium* oocysts in oysters also pose a potential health concern. Other filter-feeders such as freshwater clams and marine mussels have also been shown to accumulate *Cryptosporidium* oocysts (Graczyk et al. 1998, Chalmers et al. 1997). They may serve as similar biologic monitors for *Cryptosporidium* oocyst contamination.

CONCLUSIONS

Although the traditional classification of species based on the vertebrate classes of their hosts is largely accurate, it has greatly underestimated the diversity various *Cryptosporidium* isolates. This has presented problems in the identification of parasites in environmental samples. Molecular techniques are now available to identify species of *Cryptosporidium* and to differentiate known genotypes of *C. parvum*, and should be very useful in the investigation of clinical outbreaks of cryptosporidiosis. The performance of these techniques in the analysis of environmental samples, however, has yet to be thoroughly demonstrated. Because of the nature of environmental samples, *Cryptosporidium* isolates from various hosts must be more extensively characterized before enough data have been acquired and interpreted to instill full confidence in the method.

ACKNOWLEDGMENTS

To Joself Limor for technical assistance.

REFERENCES

Awad-El-Kariem FA, Robinson HA, Dyson DA, Evans D, Wright S, Fox MT, McDonald V 1995. Differentiation between human and animal strains of *Cryptosporidium parvum* using isoenzyme typing. *Parasitology 110*: 129-132.

Awad-El-Kariem FA, Warhurst DC, McDonald V 1994. Detection of *Cryptosporidium* oocysts using a sytem based on PCR and endonuclease restriction. *Parasitology 109*: 19-22.

Bonnin A, Fourmaux MN, Dubremetz JF, Nelson RG, Gobet P, Harly G, Buisson M, Puygauthier-Toubas D, Gabriel-Pospisil F, Naciri M, Camerlynch P 1996. Genotyping human and bovine isolates of *Cryptosporidium parvum* by polymerase chain reaction-restriction fragment length polymorphism analysis of a repetitive DNA sequence. *FEMS*

Microbiol Lett 137: 207-211.

Cai J, Collins MD, McDonald V, Thompson DE 1992. PCR cloning and nucleotide sequence determination of the 18S rRNA genes and internal transcribed spacer 1 of the protozoan parasites *Cryptosporidium parvum* and *Cryptosporidium muris*. *Biochem Biophys Acta 1131*: 317-320.

Carraway M, Widmer G, Tzipori S 1994. Genetic markers differentiate *C. parvum* isolates. *J Eukaryot Microbiol 41*: 26S.

Carraway M, Tzipori S, Widmer G 1996. Identification of genetic heterogeneity in the *Cryptosporidium parvum* ribosomal repeat. *Appl Env Microbiol 62*: 712-716.

Carraway M, Tzipori S, Widmer G 1997. A new restriction fragment length polymorphism from *Cryptosporidium parvum* identifies genetically heterogeneous parasite populations and genotypic changes following transmission from bovine to human hosts. *Infect Immun 65*: 3958-3960.

Chalmers RM, Sturdee AP, Mellors P, Nicholson V, Lawlor F, Kenny F, Timpson P 1997. *Cryptosporidium parvum* in environmental samples in the Sligo area, Republic of Ireland: a preliminary report. *Lett Appl Microbiol* 25: 380-384.

Current WL, Upton SJ, Haynes TB 1986. The life cycle of *Cryptosporidium baileyi* n. sp. (Apicomplexa, Cryptosporidiidae) infecting chickens. *J Protozool 33*: 289-296.

Fayer R, Graczyk TK, Lewis EJ, Trout JM, Farley CA 1998. Survival of infectious *Cryptosporidium parvum* oocysts in seawater and eastern oysters (*Crassostrea virginica*) in the Chesapeake bay. *Appl Environ Microbiol 64*: 1070-1074.

Fayer R, Speer CA, Dubey JP 1997. The general biology of *Cryptosporidium*, p. 1-41. In R Fayer, *Cryptosporidium and Cryptosporidiosis*, CRC Press, Boca Raton, Florida.

Graczyk TK, Fayer R, Cranfield MR, Conn DB 1998. Recovery of *waterborne Cryptosporidium parvum* oocysts by freshwater benthic clams (*Corbicula fluminea*). *Appl Environ Microbiol 64*: 427-430.

Kilani RT, Wenman WM 1994. Geographical variation in 18S rDNA gene sequence of *Cryptosporidium parvum*. *Int J Parasitol 24*: 303-306.

Leng X, Mosier DA, Oberst RD 1996. Differentiation of *Cryptosporidium parvum*, *C. muris*, and *C. baileyi* by PCR-RFLP analysis of the 18s rRNA gene. *Vet Parasitol 62*: 1-7.

Levine ND 1984. Taxonomy and review of the coccidian genus *Cryptosporidium* (Protozoa, Apicomplexa). *J Protozool 31*: 94-98.

Levine ND 1986. The taxonomy of *Sarcocystis* (Protozoa, Apicomplexa) species. *J Parasitol 72*: 372-382.

Mead JR, Humphreys RC, Sammons DW, Sterling C 1990. Isolation of isolate-specific sporozoite proteins of *Cryptosporidium parvum* by two-dimensional gel electrophoresis. *Infect Immun 58*: 2071-2075.

Morgan UM, Constantine CC, Forbes DA, Thompson RCA 1997. Differentiation between human and animal isolates of *Cryptosporidium parvum* using rDNA sequencing and direct PCR analysis. *J Parasitol 83*: 825-830.

Morgan UM, Constantine CC, O'Donoghue P, Meloni BP, O'Brien PA, Thompson RCA 1995. Molecular characterization of *Cryptosporidium* isolates from humans and other animals using random amplified polymorphic DNA analysis. *Am J Trop Med Hyg 52*: 559-564.

Nichols GL, McLauchlin J, Samuel D 1991. A technique for typing *Cryptosporidium* isolates. *J Protozool 38*: 237S-240S.

Nina JMS, McDonald V, Dyson DA, Catchpole J, Uni S, Iseki M, Chiodini PL, McAdam KPWJ 1992. Analysis of oocyst wall and sporozoite antigens from three *Cryptosporidium* species. *Infect Immun 60*: 1509-1513.

O'Donoghue PJ 1995. *Cryptosporidium* and cryptosporidiosis in man and animals. *Int J Parasitol 25*: 139-195.

Ogunkolade BW, Robinson HA, McDonald V, Webster K, Evans DA 1993. Isoenzyme variation within the genus *Cryptosporidium*. *Parasitol Res 79*: 385-388.

Ortega YR, Sheehy RR, Cama VA, Oishi KK, Sterling CR 1991. Restriction fragment length polymorphism analysis of *Cryptosporidium parvum* isolates of bovine and human origin. *J Protozool 38*: 540-541.

Peng MP, Xiao L, Freeman AR, Arrowood MJ, Escalante A, Weltman AC, Ong C, Mac Kenzie WR, Lal AA, Beard CB 1997. Genetic polymorphism among *Cryptosporidium parvum* isolates supporting two distinct transmission cycle. *sEmer Infect Dis 3*: 1-9.

Spano F, Putignani L, McLauchlin J, Casemore DP, Crisanti A 1997. PCR-RFLP analysis of the *Cryptosporidium* oocyst wall protein (COWP) gene dicriminates between *C. wrairi* and *C. parvum*, and between *C. parvum* isolates of humand animal origin. *FEMS Microbiol Lett 150*: 209-217.

Tzipori S, Angus KW, Campbell I, Gray EW 1980. *Cryptosporidium*: evidence for a single-species genus. *Infect Immun 30:* 884-886.

Tzipori S, Griffiths JK 1998. Natural history and biology of *Cryptosporidium parvum*. *Adv Parasitol 40*: 5-36.

Upton SJ, Current WL 1985. The species of *Cryptosporidium* (Apicomplexa: Cryptosporidiidae) infecting mammals. *J Parasitol 71*: 625-9.

Vasquez JR, Gooze L, Kim K, Gut J, Petersen C, Nelson RG 1996. Potential antifolate resistance determinants and genotypic variation in the bifunctional dihydrofolate reductase-thymidylate synthase gene from human and bovine isolates of *Cryptosporidium parvum*. *Mol Biochem Parasitol 79*: 153-165.

RESEARCH NOTE

Human Genetic Bi-allelic Sequences (HGBASE), a Database of Intra-genic Polymorphisms

Chandra Sarkar/+ , Flavio R Ortigão, Ulf Gyllensten*/, Anthony J Brookes****

Interactiva Biotechnologie GmbH, D-89077 Ulm, Germany *Swedish Genome Research Center **Department of Genetics and Pathology, Biomedical Center, Uppsala, Sweden

Key words: single nucleotide polymorphisms - polymorphisms - intra-genic polymorphisms - databases - bioinformatics

The Human Genome Project is providing a wealth of information about the human gene repertoire, and promises to furnish a complete genome sequence (and thereby a complete gene catalog) by the year 2005. This enormous output of data is beginning to be complemented by large scale studies designed to uncover normally occurring variations within human gene sequences. Much of this variability is very subtle, often comprises single nucleotide polymorphisms (SNPs) which are ideally compatible with a number of large scale detection procedures. SNPs will be the basis of future highly dense polymorphic marker maps, and those related to known genes can be exploited in genetic association studies aimed at defining the genetic basis of all manner of complex phenotypes, not least disorders such as mental illness, diabetes, cardiovascular disease and cancer. All indications are that 100,000-200,000 human genome SNPs will be identified within the next two years.

In light of the above developments, a database of gene based polymorphisms is obviously required. To fulfill this need we have constructed and recently released at http://hgbase.interactiva.de the HGBASE (human genic bi-allelic sequences) database of intra-genic sequence polymorphism. HGBASE is the result of a joint venture between Uppsala University Medical Genetics Department, the Swedish Genome Research Centre, and Interactiva Biotechnologie GmbH. Its primary purpose is to facilitate genotype-phenotype association studies based upon the rapidly growing number of known, gene related, single nucleotide polymorphisms (SNPs) and other intra-genic sequence variations. Furthermore, HGBASE will help towards the production of a dense SNP map of the human genome, which itself will be a valuable research tool.

HGBASE is not designed to include gene 'mutations', but instead is a catalog of intra-genic (promoter to transcription end point) sequence variants found in 'normal' individuals. Although the distinction between 'mutation' and 'variation' can be somewhat blurred, the general idea is that the content of HGBASE concerns frequently occurring 'normal polymorphisms', whether or not they are suspected to increase the risk of developing a particular phenotype. This is in contrast to 'mutant sequences' which are known to cause genetic disease. Despite its name, HGBASE contains all types of intra-genic variation and is not limited to bi-allelic polymorphisms (though these do represent most of the database content). Both functional polymorphisms (e.g. promoter and non-silent codon changes) and non-functional polymorphisms (e.g. intron sequence differences) are included. This is for two reasons. Firstly, it is often difficult to be certain about the functional consequence of a variation. Secondly, regardless of functional relevance, any intra-genic polymorphism can usually be employed as an effective surrogate marker for an unknown functional variant in an association study, due to close proximity and linkage disequilibrium.

Gene polymorphisms may be retrieved from HGBASE by using the database search facilities to query either by a text string or by a DNA sequence. Data submission to HGBASE is made simple by provision of a series of Web page data submission forms. All submitted data is made available to any other public database that wishes to download it, and continual efforts are made to access new relevant data from other databases and literature publications. The exponential growth in polymorphism discovery requires that scientists make every effort to submit their data to the HGBASE database to ensure it remains up to date. HGBASE does not claim any rights to publicly available or submitted data, instead this remains the property of the original submitter. Deposition of data into HGBASE requires only the allelic DNA sequences, the allele frequencies, the host gene name, and the intra-genic domain. Additional com-

+Corresponding author. Fax: +49-731-93579291. E-mail: sarkar@interactiva.de
Received 15 June 1998
Accepted 30 July 1998

ments, such as assay conditions, can be supplied though are not required. The submitted data is presented in HGBASE along with the submitters name and contact details to aid discussion and questioning. Database curators will subsequently enhance the submitted data by adding links to other databases, and by adding information concerning gene function, gene location, gene expression pattern, disease associations, and suggested assay formats. This 'added value' data is accessible to users following a simple registration procedure that is free to academia but for which a charge is made to industry to cover the costs of collecting and maintaining the additional data.

Selection, Recombination and History in a Parasitic Flatworm (*Echinococcus*) Inferred from Nucleotide Sequences

KL Haag, AM Araújo, B Gottstein*, A Zaha**

Departamento de Genética, Universidade Federal do Rio Grande do Sul, Caixa Postal 15053, 91501-970 Porto Alegre, RS, Brasil *Institute of Parasitology, University of Berne, Laengass Strasse 122, PO Box 8466, Berne, CH-3001, Switzerland **Departamento de Biotecnologia, Universidade Federal do Rio Grande do Sul, Caixa Postal 15005, 91501-970 Porto Alegre, RS, Brasil

Three species of flatworms from the genus Echinococcus *(*E. granulosus, E. multilocularis *and* E. vogeli*) and four strains of* E. granulosus *(cattle, horse, pig and sheep strains) were analysed by the PCR-SSCP method followed by sequencing, using as targets two non-coding and two coding (one nuclear and one mitochondrial) genomic regions. The sequencing data was used to evaluate hypothesis about the parasite breeding system and the causes of genetic diversification. The calculated recombination parameters suggested that cross-fertilisation was rare in the history of the group. However, the relative rates of substitution in the coding sequences showed that positive selection (instead of purifying selection) drove the evolution of an elastase and neutrophil chemotaxis inhibitor gene (AgB/1). The phylogenetic analyses revealed several ambiguities, indicating that the taxonomic status of the* E. granulosus *horse strain should be revised.*

Key words: *Echinococcus* - parasites - recombination - SSCP - sequencing - phylogeny

Several new insights about the evolution of helminth parasites came out during the last years. *Echinococcus*, a parasite that causes one of the most important and widespread zoonoses, the hydatid disease, is included in this group. The small flatworm uses herbivores as intermediate hosts and carnivores as final hosts. The adult is hermaphrodite and the larval stage (metacestode) is amplified by asexual reproduction.

Four species within the genus are recognised: *E. vogeli* and *E. oligarthrus*, which occur in the neotropical region, *E. multilocularis*, that has an holartic geographic range and *E. granulosus*, that is world-wide distributed. Due to a low intermediate host specificity, *E. granulosus* has been subdivided in several strains, according to the host species used, or to the geographic range of the biological cycle. Some of the evolutionary questions concerning *Echinococcus* are: (1) is the adult mainly self- or cross-fertilising? (2) how do the strains within a species differentiate? (3) what is the true taxonomic status of these strains?

The first question relates to the second one: depending on the breeding system, only one of two modes of strain differentiation can occur. If individual parasites would be mainly selfers (Smyth & Smyth 1964), purifying (negative) selection would quickly eliminate the non-adaptive mutations, due to increased homozygosis. In addition, selfing would lead to a high rate of linkage disequilibrium within parasite populations. In this situation, the genome would be selected as a whole, and not in pieces of recombining DNA. If, on the other hand, populations would undergo outcrossing (Rausch 1967, 1985), free recombination would allow genes to be selected as individual units, and

This work was supported by the Swiss National Science Foundation (project n° 31-45575.95), PADCT/CNPq (Proc. 620081/95-3), EEC (DG XII CI 10284-0), the "Jubiläumsstiftung der Schweitzerischen Lebensversicherungs- und Rentenanstalt für Volksgesundheit und Medizinische Forschung" and the "Sandoz-Stiftung zur Förderung der medizinisch-biologischen Wissenschaften".

This paper reports on research conducted by Karen Luisa Haag as part of her PhD thesis on strain characterisation, genetic variability and breeding systems of *Echinococcus*. It is a result of a collaborative work between the Centro de Biotecnologia (Universidade Federal do Rio Grande do Sul, Brazil) and the Institute of Parasitology (University of Berne, Switzerland). Arnaldo Zaha works primarily with gene organisation and control in *E. granulosus*, Aldo Mellender de Araújo works with evolutionary ecology on a variety of organisms, but mainly insects, and Bruno Gottstein is dealing with molecular aspects of host-parasite interactions in *E. multilocularis*.
+Corresponding author. Fax: +55-51-319.2011. E-mail: haag@dna.cbiot.ufrgs.br
Received 15 June 1998
Accepted 30 July 1998

each genomic sequence would be able to respond singularly to the positive and/or negative selection imposed by the host.

It has also been argued (Thompson et al. 1995, Lymbery & Thompson 1996) that the degree of genetic differentiation of some strains is larger than expected for conspecific groups. Furthermore, if *Echinococcus* is an obligatory selfer, the biological species concept cannot be used to solve the problem (Lymbery 1992, Lymbery & Thompson 1996). In the present study we used the nucleotide sequencing of two coding and two non-coding regions of *Echinococcus* genome to try to elucidate some of the questions above. If parasite populations would have undergone outcrossing during their evolutionary history, we would expect to find recombination among sequences. Additionally, by assessing relative rates of substitution in coding and non-coding regions, it would be possible to evaluate the occurrence of positive and/or negative selection. Finally, genetic distances estimated from those sequences could help to decide whether or not some of the *E. granulosus* strains should be regarded as different species.

MATERIALS AND METHODS

Molecular analyses - Thirty three *E. multilocularis* isolates from different continents (Asia, Europe and North America), hosts (foxes, humans and rodents) and life cycle stages, as well as 110 *E. granulosus* metacestode isolates from different geographic regions (Australia, Europe and Southern Brazil) and strains (bovine, equine, ovine and swine) and one *E. vogeli* isolate were used for genomic DNA extraction and further analyses. DNA extraction was done by standard procedures (McManus & Simpson 1985).

For each isolate, four different targets were amplified by PCR, using primers specific for *Echinococcus* DNA (see procedures in Haag et al. 1997). Two of them were partial intron sequences from an actin gene (ActII - 266 bp) and from an homeobox containing gene (Hbx2 - 331 bp). The other two were coding regions: a partial sequence of a neotrophil chemotaxis inhibitor nuclear gene (AgB/1 - 101 bp) and another partial sequence of the mitochondrial NADH dehydrogenase 1 gene (ND1 - 141 bp).

The nucleotide variation within the PCR products obtained for the four targets was screened by the PCR-SSCP method (see procedures in Haag et al. 1997). Subsequently, two isolates from each SSCP pattern (except in the case of *E. vogeli*) were chosen for direct fluorescence sequencing. For this, the single stranded DNA bands were cut out from the fresh silver-stained SSCP gels, washed and eluted. One ml of the eluted single strands was used for re-amplification with the corresponding primers. These re-amplification products and their respective primers were used for sequencing.

Statistic and phylogenetic analyses - Sequences were aligned by eye (Fig. 1) and the molecular diversity parameters, recombination rates and relative rates of synonymous/non-synonymous substitutions (Ka/Ks) were estimated using DnaSP version 2.0 (Rozas & Rozas 1997). The recombination parameter (C) is calculated based on the average number of nucleotide differences between pairs of sequences (Hudson 1987) and a minimum number of recombination events in the history of the sample (RM) is obtained using a four-gamete test (Hudson & Kaplan 1985).

The genetic distances as well as the neighbour-joining (NJ) trees were estimated with MEGA version 1.0 (Kumar et al. 1993). The parsimony trees were constructed using DNA Penny in Phylip version 3.5c (Felsenstein 1993). For the NJ phylogenetic analysis we used a gamma distance (Kimura 2-parameter model) with gamma parameter a=1. In the parsimony analysis we made a branch-and-bound search to find all most parsimonious trees. Both kinds of trees were constructed using *E. vogeli* as outgroup.

RESULTS

The degree of allele polymorphism found within *E. multilocularis* and within strains of *E. granulosus* was low, as shown in our previous studies (Haag et al. 1997, 1998). Indeed, only one transversion and a single base deletion in the Hbx2 intron occurred among isolates of *E. multilocularis* (Haag et al. 1997). Within the cattle, horse, pig and sheep strains of *E. granulosus* no allele polymorphism was found in the four coding and non-coding loci analysed in the present study.

For this reason, further analyses were done considering the most common variant of *E. multilocularis*, the sequences of the four *E. granulosus* strains and those obtained for the *E. vogeli* isolate {GenBank assession numbers are: AF003748, AF003749, AF003750, AF024661 and AF024662 (Act II); X66818, AF003976, AF003977, AF024663 and AF024664 (Hbx 2); Z26481, Z26482, Z26483, Z26336 and AF024665 (AgB/1); U65748 [ND1 - authors did not provide information about variant sequences published by Bowles and McManus (1993)]}. The molecular diversity parameters estimated from this data set are shown in Table I. The most variable locus was the mitochondrial ND1. Surprisingly, one of the introns (Hbx2) was shown to be very conserved among the referred strains and species, and the AgB/1 nuclear coding region had as much variability as the Act II intron.

A

```
                  10                  30                  50
                   '         '         '         '         '         '

sheep     TCGTCCAAGACATCAGGTTAGTTGGATAGGTAGGCAGTGTTTCAGCCGCACCGGAACTGG
cattle    .............................................T..........G...
pig       .............................................T..........G...
horse     ............................................................
multiloc  ............................................................
vogeli    .......................A...G........A.......................

                  70                  90                 110
                   '         '         '         '         '         '

sheep     TACCAACTAGTGGACCAATTTTCTCAAATAAGAGACAGAAATGGTTTGCTTTCATGCACT
cattle    ...........C.........C......A............CA.......C.........
pig       ...........C.........C......A............CA.......C.........
horse     .T.........C.........CTG..T..............CA.......C.........
multiloc  .T.........C...T.....CTG..T..........A...CA.......C.........
vogeli    .T.........C.........C....T..............CA.......C.........

                 130                 150                 170
                   '         '         '         '         '         '

sheep     AAATGTATGGTGAAGAAGTCGGCTTTTCATCTAACTAGATAGGCATGATTAGTGTGGAGA
cattle    .......................................G....................
pig       ......................T.....................................
horse     .................................G...........A..............
multiloc  ..............................................A.............
vogeli    ....A.........................................A.............

                 190                 210                 230
                   '         '         '         '         '         '

sheep     TCAAGTGCTCTCTTGTAGAGTCGCCATCTGAGGGCAGTCTTTCTATTTTCGCCCTGTGAC
cattle    .......................T...T................................
pig       ...........................T.........GG.C...................
horse     .................A.........T................................
multiloc  ............................................................
vogeli    .................................................T..GGG.....

                 250
                   '         '

sheep     AACGTACCTATTCCGAAATAATCTTT
cattle    .......................A..
pig       .......................A..
horse     .......................A..
multiloc  ..........................
vogeli    ................G.........
```

Fig. 1-A: nucleotide sequence alignments of the ActII intron for the *Echinococcus granulosus* sheep, cattle, pig and horse strains as well as for *E. multilocularis* and *E. vogeli*.

B

```
                    10                  30                  50
                     '         '         '         '         '         '
sheep       CGTCTTAGAAGAGCGATTTGATCGACAAAAGTACCTCAGCAGTGCTGAACGCGCCGAGAT
cattle/pig  ............................................................
horse       ............................................................
multiloc    ............................................................
vogeli      ....................................A...T................C..

                    70                  90                 110
                     '         '         '         '         '         '
sheep       GTCACGAGACCTGGGGCTCTCTGAAACCCAGGTATGTCACAGCCGATGTCATTAAACATG
cattle/pig  ..............................................C.............
horse       ..............................................C.............
multiloc    ..............................................C.............
vogeli      .....A.A......................................C.............

                   130                 150                 170
                     '         '         '         '         '         '
sheep       GGAAGGGGTGAGAGTAGTTGGAGCGTCACGAAGTGCCAAATTGGGCGCTTGTCAAGCTGC
cattle/pig  ................T...........................................
horse       ............................................................
multiloc    ............................................................
vogeli      ..................CA........................................

                   190                 210                 230
                     '         '         '         '         '         '
sheep       GCCTTTATAACTGTTGAGTGCATCATCACCCATAAAAAATTGGGAGAGAGGGGGGCGGGA
cattle/pig  ............................................................
horse       ...........................................-................
multiloc    ............................................................
vogeli      .............................................T..............

                   250                 270                 290
                     '         '         '         '         '         '
sheep       GCGGGTCAAAAGGGTCATCACGGCTCATGCATTAGTAAGATCGTAAAAGGCATGCCTCTA
cattle/pig  ............................................................
horse       ............................................................
multiloc    ....................T.......................................
vogeli      ...................GGT.......A..................C...........

                   310                 330
                     '         '         '
sheep       ATTATGACCCCCACCACTAGGTGAAAATATG
cattle/pig  ...............................
horse       ...............................
multiloc    .........T.....................
vogeli      ...............................
```

Fig. 1-B: nucleotide sequence alignments of the Hbx2 intron for the *Echinococcus granulosus* sheep, cattle, pig and horse strains as well as for *E. multilocularis* and *E. vogeli*.

C

```
                   10                  30                  50
                    '         '         '         '         '         '

sheep      CTGGTTGGGGTGGTTACAACAATTATTCATTTTTAAGGTCGGTTCGATGTGCTTTTGGAT
cattle     .A.........A....T..T.G......T...............................
pig        ...........A.......T.G......T...............................
horse      .G.........A....T..T.....C........G.....A...................
multiloc   ...........A.......T..A.....G........T..T.................G.
vogeli     .......A...A.......T..C........C..G.....AA....G...........G.

                   70                  90                 110
                    '         '         '         '         '         '

sheep      CTGTTAGGTTTGAGGCTTGTTTTATGTGTGTGGTGATTTTTTGTGCTTTGTGTAGTTGTA
cattle     ..................................T..................T......
pig        ..................................T........C.....A..CT.....G
horse      ............................C........................T....C.
multiloc   .............A....................A.........T........TAC....
vogeli     ..................................A......C..........CT......

                  130
                    '         '

sheep      GGTATAATTTAATTGATTTTT
cattle     .....................
pig        .....................
horse      ..........GG.........
multiloc   ..........G..........
vogeli     ..........GG.........
```

Fig. 1-C: nucleotide sequence alignments of the mitochondrial ND1 for the *Echinococcus granulosus* sheep, cattle, pig and horse strains as well as for *E. multilocularis* and *E. vogeli*.

D

```
10                     30                  50
                        '         '         '         '         '         '

sheep          AGTGGTTGACCTCTTAAAGGAACTGGAAGAAGTGTTCCAGTTGTTGAGGAAGAAGCTACG
cattle/pig     ............................................................
horse          ......................................G.....................
multiloc       .T..........A...........................A....A..............
vogeli         ..........A.G...............................................

                       70                  90
                        '         '         '         '

sheep          CATGGCACTCAGGTCCCACCTCAGAGGGTTGATTGCTGAAGG
cattle/pig     ..C......T.A..............A....G..........
horse          ..C........A..............A....G..........
multiloc       ..C........A..............A....G..........
vogeli         ..C......T.A...........A.AA....G..........
```

Fig. 1-D: nucleotide sequence alignments of the AgB/1 for the *Echinococcus granulosus* sheep, cattle, pig and horse strains as well as for *E multilocularis* and *E. vogeli*.

TABLE I

Nucleotide diversity (π), theta (θ), average number of nucleotide differences (k), number of polymorphic sites (S) and total number of sites (T) of the four non-coding (Act II and Hbx 2) and coding (AgB/1 and ND1) sequences analysed in this study

PARAMETER[a]	Act II	Hbx 2	AgB/1	ND1
π	0.0524	0.0204	0.0559	0.0964
	(0.0001)[b]	(0.0001)	(0.0001)	(0.0002)
θ	0.0576	0.0233	0.0618	0.0963
	(0.0008)	(0.0002)	(0.0011)	(0.0023)
k	13.93	6.70	5.70	13.60
S	35	16	13	31
T	266	329	102	141
S/T	0.1316	0.0486	0.1274	0.2198

a: Nei 1987; *b:* Numbers in parentheses are standard deviations.

The recombination parameter (C=4Nc, where c is the recombination rate) among the nuclear sequences was equal to 34.2 (per gene) and 0.0518 (between adjacent sites). The minimum number of recombination events occurring in the history of that sample of sequences was estimated do be Rm=2. Additionally, the relative rates of synonymous and non-synonymous substitutions calculated for the two coding regions showed that, compared to the mitochondrial ND1, the rates of non-synonymous substitutions within AgB/1 were very high (Table II).

As the results of the NJ and parsimony analyses were very similar, we decided to concentrate on the later. A phylogeny obtained by analysing all loci together is shown in Fig. 2. The topology of that tree is in accordance with others, obtained using a larger number of OTUs and other helminths as outgroups (Lymbery 1995). However, the phylogenies constructed for each sequence separately were not congruent. First, most sequences did not provide a single most parsimonious tree: the Hbx2 intron resulted in 15, ND1 in 2 and AgB1 in three equally parsimonious topologies. Second,

TABLE II

Relative rates of non-synonymous and synonymous (Ka/Ks) substitutions within ND1 (above diagonal) and AgB1 (below diagonal) coding sequences among the *Echinococcus granulosus* strains, *E. multilocularis* (EM) and *E. volgeli* (EV)

	Sheep	Cattle	Pig	Horse	EM	EV
Sheep		0.20	0.17	0.09	0.13	0.10
Cattle	1.22		0.07	0.06	0.08	0.07
Pig	1.22	0.00		0.05	0.08	0.09
Horse	*	0.31	0.31		0.07	0.04
EM	0.88	0.18	0.18	0.45		0.14
EV	1.48	2.09	2.09	0.90	0.61	

* indeterminacy

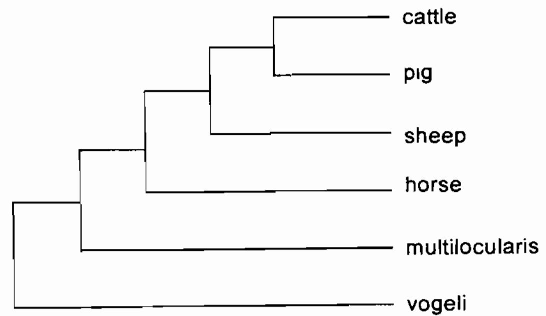


Fig. 2: maximum parsimony phylogenetic tree of *Echinococcus* strains and species obtained using the four coding and non-coding sequences. The tree requires 113 steps (for details, see Materials and Methods).

ambiguities were found regarding the position of the horse strain: in some instances it is grouped together with the *E. granulosus* strains, and in others it splits before.

A striking result obtained by the genetic distance calculations (Table III) was the high similarity between the cattle and the pig strains. As expected, *E. vogeli* is the most distant group in relation to all other analysed OTUs. The distance values among the other *E. granulosus* strains and between each strain and *E. multilocularis* were quite similar.

DISCUSSION

Previous studies (Lymbery et al. 1997) concluded that cross-fertilisation occurs within *E. granulosus* populations. However, there were also good evidences that outcrossing is not the predominant mating system, since most loci analysed showed monomorphism within strains or large deficiencies of heterozygotes (Lymbery & Thompson 1988, Lymbery et al. 1990, 1997). The results obtained in the present study support those previ-

TABLE III

Jukes-Cantor genetic distances (above diagonal) and their standard deviations (below diagonal) among the *Echinococcus granulosus* strains, *E. multilocularis* (EM) and *E. volgeli* (EV), based on the nuleotide sequences of the four coding and non-coding loci

	Sheep	Cattle	Pig	Horse	EM	EV
Sheep		0.0329	0.0379	0.0392	0.0455	0.0700
Cattle	0.0064		0.0145	0.0317	0.0442	0.0674
Pig	0.0069	0.0042		0.0405	0.0493	0.0700
Horse	0.0070	0.0062	0.0071		0.0392	0.0622
EM	0.0075	0.0074	0.0079	0.0070		0.0635
EV	0.0095	0.0093	0.0095	0.0089	0.0090	

ous findings, suggesting that recombination within the nuclear sequences occurred at least twice during the evolution of the genus. Although the coding and non-coding regions tested here were short, the lack of phylogenetic congruence among the trees constructed for each locus separately could also be due to recombination.

Another explanation for those incongruences is that selection acted independently on each sequence, but this argument could be used only for the coding regions. Indeed, we showed that positive selection did act during the evolution of the AgB/1 gene: most nucleotide replacements found by pairwise comparisons of the sequences were non-synonymous, and the relative rates of non-silent/silent substitutions (Ka/Ks) were greater than one in six out of fifteen comparisons.

Selection was also used to explain the high frequency of heterozygotes found for variant regulatory sequences in populations of *E. granulosus* from the sheep strain. Taken together, all those findings indicate that *Echinococcus* is not an evolutionary dead-end, unable to adapt quickly enough to changing environmental conditions. Nevertheless, it seems that a balance between cross and self-fertilisation was the best solution found by the parasite to keep evolving. It seems that the recombination rates cannot be neither too high, breaking down coadapted gene complexes, nor too low, hindering adaptive changes.

Moreover, the estimated phylogenetic distances and the trees of *Echinococcus* species and strains are in agreement with those reported by Lymbery (1995). The results show that the phylogenetic position of the *E. granulosus* horse strain is ambiguous. For this reason, we agree with the proposal of a taxonomic revision of the genus, based not only on a molecular phylogenetic approach including a larger number of OTUs, but also on other comparative biological data.

REFERENCES

Bowles J, McManus DP 1993. NADH dehydrogenase 1 gene sequences compared for species and strains of the genus *Echinococcus*. *Internl J Parasitol 23:* 969-972.

Felsenstein J 1993. *Phylip (Phylogeny Inference Package)* version 3.5c. Distributed by the author. Department of Genetics, University of Washington, Seattle.

Haag KL, Zaha A, Araújo AM, Gottstein B 1997. Reduced genetic variability in coding and non-coding regions of *Echinococcus muitilocularis* genome. *Parasitology 115:* 521-530.

Haag KL, Araújo AM, Gottstein B, Siles-Lucas M, Thompson RCA, Zaha A 1998. Breeding system in *Echinococcus granulosus* (Cestoda; Taeniidae); selfing or outcrossing? *Parasitology* (in press).

Hudson RR 1987. Estimating the recombination parameter of a finite population model without selection. *Gen Res 50*: 245-250.

Hudson RR, Kaplan NL 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics 111*: 147-164.

Kumar S, Tamura K, Nei M 1993. *Mega: Molecular Evolutionary Genetics Analysis,* version 1.0. The Pensylvania State University, University Park, PA 16802.

Lymbery AJ 1992. Interbreeding, monophyly and the genetic yardstick: species concepts in parasites. *Parasitol Today 8*: 208-211.

Lymbery AJ 1995. Genetic diversity, genetic differentiation and speciation in the genus *Echinococcus* Rudolphi 1801, p. 51-88. In RCA Thompson, AJ Lymbery (eds), *Echinococcus and Hydatid Disease,* Cab International, Wallingford.

Lymbery AJ, Thompson RCA 1988. Electrophoretic analysis of genetic variation in *Echinococcus granulosus* from domestic hosts in Australia. *Intern J Parasitol 18*: 803-811.

Lymbery AJ, Thompson RCA 1996. Species of *Echinococcus*: pattern and process. *Parasitol Today 12*: 486-491.

Lymbery AJ, Constantine CC, Thompson RCA 1997. Self-fertilization without genomic or population

structuring in a parasitic tapeworm. *Evolution 51*: 289-294.

Lymbery AJ, Thompson RCA, Hobbs RP 1990. Genetic diversity and genetic differentiation in *Echinococcus granulosus* (Batsch, 1786) from domestic and sylvatic hosts on the mainland of Australia. *Parasitology 101*: 283-289.

McManus DP, Simpson AJG 1985. Identification of the *Echinococcus* (hydatid disease) organisms using cloned DNA markers. *Mol Biochem Parasitol 17*: 171-178.

Nei M 1987. *Molecular Evolutionary Genetics*, Columbia University Press, New York, 512 pp.

Rausch RL 1967. A consideration of intraspecific categories in the genus *Echinococcus* Rudolphi,1801 (Cestoda: Taeniidae). *J Parasitol 53*: 484-491.

Rausch RL 1985. Parasitology: retrospect and prospect. *J Parasitol 71*: 139-151.

Rozas J, Rozas R 1997. DnaSP version 2.0: a novel software package for extensive molecular population genetics analysis. *Computation Applicated to Biosciences 13*: 307-311.

Smyth JD, Smyth MM 1964. Natural and experimental hosts of *Echinococcus granulosus* and *E. multilocularis*, with comments on the genetics of speciation in the genus *Echinococcus*. *Parasitology 54*: 493-514.

Thompson RCA, Lymbery AJ, Constantine CC 1995. Variation in *Echinococcus*: towards a taxonomic revision of the genus. *Advanc Parasitol 35*: 146-176.

RESEARCH NOTE

Surveillance Using Molecular Tools: Examples from Brazil

Ana Carolina Paulo Vicente/+, Hooman Momen*

Departamento de Genética *Departamento de Bioquímica e Biologia Molecular, Instituto Oswaldo Cruz, Av. Brasil 4365, 21045-900 Rio de Janeiro, RJ, Brasil

Key words: surveillance - HIV2 - *Vibrio* - cholera - *Leishmania* - *Escherichia coli*

Brazil presents particular problems for surveillance of infectious diseases. These include its continental size, uneven distribution of resources, difficulty of communication and access in some of the more remote areas, as well as large areas covered by tropical rain-forests. Surveillance for infectious diseases in Brazil has traditionally been carried out in a passive manner by government authorities or as individual initiatives. Most effort has been directed in the collecting and tabulating of data on notifiable infections. A limited amount of laboratory support has been available for the isolation and identification of the etiological agents. More recently molecular methods have been introduced in the analysis of these data.

Here we present some practical examples of the use of different molecular tools for diagnosis and in the analysis of infectious diseases and in epidemiological monitoring of outbreaks. The examples are taken from work carried out in our Institute.

HIV-2, the second AIDS-causing virus, was originally identified and found to be quite common in West Africa. With a more restricted geographic spread than HIV-1, this virus has also been isolated in countries with socioeconomic links to West Africa (R Marlink 1996 *AIDS 10*:689-699). Some early reports analyzing the seroprevalence of HIV-1 and HIV-2 were contradictory about the presence of HIV-2 in Brazil (L Oyafuso et al. 1989 *New Engl J Med 320*: 953-958, RM Hendry et al. *J Acq Imm Def Synd 4*: 623-627). At that time they concluded that there was some cross-reactivity between HIV-2 and HIV-1 which resulted in misinterpretation. Using polymerase chain reaction (PCR) and specific internal probes to HIV-2, Pieniazek et al. (1991 *AIDS 5*: 1293-1299), identified mixed HIV-1/HIV-2 infections in Rio de Janeiro, Brazil. In order to validate the World Health Organization strategy for HIV testing, sera from 9,885 blood donors from São Paulo were screened by HIV enzyme-linked immunosorbent assays (ELISA) and Western blot and the results did not support the evidence of HIV-2 circulation in Brazil (MB Carvalho et al. 1996 *AIDS 10:* 1135-1140).

We have applied molecular tools in surveillance for the detection of HIV-2 in HIV-1 positive samples (possible dual infections) as well as in samples with undetermined Western blots. More than 200 samples from different parts of the country were screened for the presence of HIV-2 proviral DNA using nested PCR targeting the long terminal repeat (LTR), protease and *gag* regions. In three samples only PCR products corresponding to the LTR region were amplified. These products were sequenced and the nucleotide sequence was different from that of HIV-2 LTR. They matched with human genome sequence, probably a rare allele present in few people. At present we have failed to detect and confirm the circulation of HIV-2 in Brazil. We have shown that the use of LTR diagnostic primers to HIV-2 has to be carefully analyzed.

Vibrio cholerae occurs naturally in aquatic systems where it may constitute part of the normal microflora of zooplankton and larger animals. *V. cholerae* is a heterogeneous species with more than 140 serotypes, only a few of which are associated with biotypes causing human cholera and epidemics. The ongoing cholera pandemic (7th) is caused by the El Tor biotype, serotype O1. In 1991 cholera re-emerged in Brazil after being absent for a century, the previous pandemic involved *V. cholerae* classical biotype. The present situation is different in that not only is there a new *V. cholerae* biotype, but also there is now detailed knowledge about the bacterial virulence factors determining this disease and the molecular tools available for characterization of the isolates. In 1993 a new *V. cholerae* strain was identified in the State of Amazonas during surveillance using AP-PCR for molecular characterization of cholera vibrios. The *V. cholerae amazonia* variant is of the O1 serotype; it has distinct multilocus enzyme electrophoresis and AP-PCR profiles from other pathogenic O1 *V. cholerae.* About 50 isolates have been made from cases of diarrhea in the upper Amazon

+Corresponding author. Fax: +55-21-260.4282. E-mail: anapaulo@gene.dbbm.fiocruz.br
Received 15 June 1998
Accepted 30 July 1998

(Solimões) River. The microbe apparently does not harbor any of the well known virulence associated genes (e.g. the toxin gene casette and the major colonization factor, TCP); however some isolates present a cytotoxic effect for Y-1 cells (A Coelho et al. 1995 *J Clin Microbiol 33*: 114-118).

Since 1997, in Amazonas, all cholera notification is based on clinical and epidemiological diagnosis. In trying to identify cholera vibrio in apparent outbreaks of cholera occuring in São Paulo de Olivença, Juruá and Envira villages, we applied PCR - target specific to genes associated with *V. cholerae* El Tor (cholera toxin/CT and toxin coregulated pilus/TCP) and *V. cholerae amazonia* (regulatory gene/*toxR*). The results were negative but using PCR - target specific to genes associated with *Escherichia coli* enterotoxigenic (termo-label toxin / LT and termo-stable toxin / ST) (NG Tornieporth et al. 1995 *J Clin Microbiol 33*: 1371-1374) we were able to identify and characterize this bacteria and thus able to demonstrate that these acute diarrhea outbreaks, clinically very close to cholera symptoms, were not associated with any *V. cholerae*.

In *Leishmania*, a numerical zymotaxonomic study of New World *Leishmania* was carried out (E Cupolillo et al. 1994 *Am J Trop Med Hyg 50*: 296-311). The analysis involved the use of phenetic, cladistic and ordination techniques on enzyme electrophoresis data from more than 250 isolates of *Leishmania*. This study together with later work has revealed a rich diversity among isolates from the New World at both organismal and molecular levels. This diversity has provided numerous opportunities to probe questions concerning parasite evolution and biology, as well as their role in human disease. In many localities, more than one *Leishmania* species co-exists with overlapping animal hosts and vectors, as well as other pathogens. In collaboration with a number of different research groups we have studied aspects of the epidemiology of leishmaniasis in various countries of Latin America and in different regions of Brazil, in addition we have been interested in determining the autochtonous origin of certain *Leishmania* species found in the New World (H Momen et al. 1993 *Biol Res 26*: 249-255).

In most of these examples molecular identification of the etiological agents was followed by genetic analysis. The results were then forwarded to the relevant control agencies, usually the FNS (Fundação Nacional de Saúde).

INSTRUCTIONS TO AUTHORS

The ***Memórias do Instituto Oswaldo Cruz*** is published by the Oswaldo Cruz Institute and welcomes original contributions from research scientists throughout the world. The journal publishes original research in the fields of parasitology (protozoology, helminthology, entomology and malacology), microbiology (virology, bacteriology and mycology), tropical medicine (pathology, epidemiology and clinical studies) as well as basic studies in biochemistry, immunology, molecular and cell biology, physiology, and genetics related to these fields. Short communications in the form of Research Notes are also considered.

Manuscripts will be refereed critically by at least two reviewers; acceptance will be based on scientific content and presentation of the material.

Manuscript and figures should be sent in quadruplicate to the Editorial Office. The manuscript should be produced using a standard word processing sofware compabtible with DOS or Windows and should be types (font size, 12) double-spaced, and should be arranged in the following order: running title, title, authors' names, department or laboratory where the work was done name and address of the institution, summary, key words, introduction, materials and methods, results, discussion, acknowledgments (if any), and references. Sponsorships should be mentioned as a footnote on the first page.

Alternatively, manuscripts may be submitted following the "Uniform Requirements for Manuscripts submitted to Biomedical Journals" produced by the International Committee of Medical Journal Editors, also known as the Vancouver Style (Annals of Internal Medicine 1997; 126: 36-47).

For complete instructions, the authors should consult the home-page of the ***Memorias*** (http://www.pobox.com/~memorias), or the first issue of each volume, or contact the ***Memorias*** Office by phone (55-021-598.4335), fax (55-21-280.5048), or E-mail (memorias@pobox.com) for further information.

ERRATA

Vol. 93(1) Jan./Feb. 1998

p. 109- second paragraph- should read

R. Killick-Kendrick et al. (1994 *Ann. Trop. Med. Parasitol 88*: 183-196) demonstrated the importance of observation of the base of the spermathecal ducts for the identification of Kenyan *Phlebotomus (Larroussius)* spp. and described a technique, referred briefly by Léger et al. (1983 *Ann Paras Hum Comp 58*: 611-623), for the dissection of the female abdomen. With the insects in Berlese's fluid, they separated the terminal part, using entomological pins (size 00) attached to small wooden sticks, and covered it with a coverslip.

MEMÓRIAS DO INSTITUTO OSWALDO CRUZ

VOL. 93(5) CONTENTS SEP./OCT. 1998

3rd INTERNATIONAL MEETING ON MOLECULAR EPIDEMIOLOGY AND EVOLUTIONARY GENETICS OF INFECTIOUS DISEASES - MEEGID-3

Foreword - *Hooman Momen, Altaf A Lal, Michel Tibayrenc* 565

Plenary Lectures

Evolutionary Control of Infectious Disease: Prospects for Vectorborne and Waterborne Pathogens - *Paul W Ewald, Jeremy B Sussman, Matthew T Distler, Camila Libel, Wahid P Chammas, Victor J Dirita, Carlos André Salles, Ana Carolina Vicente, Ingrid Heitmann, Felipe Cabello* 567

Integrated Genetic Epidemiology of Infectious Diseases: The Chagas Model - *Michel Tibayrenc* 577

Molecular Epidemiologic Typing Systems of Bacterial Pathogens: Current Issues and Perpectives - *Marc J Struelens* 581

Bacterial Diseases

Vancomycin-resistant Enterococci in Intensive Care Hospital Settings - *Daren J Austin, Marc J M Bonten* .. 587

Molecular Genetic Analysis of Multi-drug Resistance in Indian Isolates of *Mycobacterium tuberculosis* - *Noman Siddiqi, Md. Shamim, NK Jain, Ashok Rattan, Amol Amin, VM Katoch, SK Sharma, Seyed E Hasnain* 589

Molecular Basis of Ribotype Variation in the Seventh Pandemic Clone and its O139 Variant of *Vibrio cholerae* - *Ruiting Lan, Peter R Reeves* 595

The Amazonia Variant of *Vibrio cholerae*: Molecular Identification and Study of Virulence Genes - *MAS Baptista, JRC Andrade, ACP Vicente, CA Salles, A Coelho* 601

Viral Diseases

Molecular Epidemiology and Emergence of Rift Valley Fever - *AA Sall, PMA Zanotto, P Vialat, OK Sène, M Bouloy* 609

Molecular Epidemiology of Human Polyomavirus JC in the Biaka Pygmies and Bantu of Central Africa - *Sylvester C Chima, Caroline F Ryschkewitsch, Gerald L Stoner* 615

Molecular Epidemiology of Den-2 Virus In Brazil - *MP Miagostovich, RMR Nogueira, HG Schatzmayr, RS Lanciotti* **(Research Note)** 625

Malaria

Antimalarial Drug Resistance: Surveillance and Molecular Methods for National Malaria Control Programmes - *Umberto D'Alessandro* 627

Allelic Diversity at the *Merozoite Surface Protein-1* (MSP-1) Locus in Natural *Plasmodium falciparum* Populations: a Brief Overview - *Marcelo U Ferreira, Osamu Kaneko, Masatsugu Kimura, Qing Liu, Fumihiko Kawamoto, Kazuyuki Tanabe* 631

Evaluation of DNA Recombinant Methodologies for the Diagnosis of *Plasmodium falciparum* and their Comparison with the Microscopy Assay - *L Urdaneta, P Guevara, JL Ramirez* 639

Systematics and Population Level Analysis of *Anopheles darlingi* - *JE Conn* 647

Anopheline Species Complexes in Brazil. Current Knowledge of Those Related to Malaria Transmission - *Maria Goreti Rosa-Freitas, Ricardo Lourenço-de-Oliveira, Carlos José de Carvalho-Pinto, Carmen Flores-Mendoza, Teresa Fernandes Silva-do-Nascimento* 651

Trypanosomatids

Implications of a Neotropical Origin of the Genus *Leishmania* - *Harry Noyes* 657

Genetic Diversity in Natural Populations of New World *Leishmania* - *Elisa Cupolillo, Hooman Momen, Gabriel Grimaldi Jr* 663

The Evolution of Trypanosomes Infecting Humans and Primates - *Jamie Stevens, Harry Noyes, Wendy Gibson* 669

Genetic Data Showing Evolutionary Links between *Leishmania* and *Endotrypanum* - *Elisa Cupolillo, Luiza OR Pereira, Octávio Fernandes, Marcos P Catanho, Júlio C Pereira, Enrique Medina-Acosta, Gabriel Grimaldi Jr* 677

Others

A Study of *Cryptosporidium parvum* Genotypes and Population Structure - *G Widmer, L Tchack, F Spano, S Tzipori* 685

Species and Strain-specific Typing of *Cryptosporidium* Parasites in Clinical and Environmental Samples - *Lihua Xiao, Irshad Sulaiman, Ronald Fayer, Altaf A Lal* 687

Human Genetic Bi-allelic Sequences (HGBASE), a Database of Intra-genic Polymorphisms - *Chandra Sarkar, Flavio R Ortigão, Ulf Gyllensten, Anthony J Brookes* **(Research Note)** 693

Selection, Recombination and History in a Parasitic Flatworm (*Echinococcus*) Inferred from Nucleotide Sequences - *KL Haag, AM Araújo, B Gottstein, A Zaha* .. 695

Surveillance Using Molecular Tools: Examples from Brazil - *Ana Carolina Paulo Vicente, Hooman Momen* **(Research Note)** 703