Evidence for a Neotropical Origin of *Leishmania*

HA Noyes, DA Morrison*, ML Chance, JT Ellis*/+**

School of Biological Sciences, University of Liverpool, Liverpool, L69 7ZD, UK *Department of Cell and Molecular Biology, University of Technology, Sydney, Westbourne St., Gore Hill, NSW 2065, Australia **Department of Parasitology, Liverpool School of Tropical Medicine, Liverpool, L3 5QA, UK

Contradictory biogeographic hypotheses for either a Neotropical or a Palaearctic origin of the genus Leishmania *have been proposed. Hypotheses constructed on the basis of biogeographic data must be tested against an independent dataset and cannot be supported by biogeographic data alone. In the absence of a fossil record for the* Leishmania *these two hypotheses were tested against a combined dataset of sequences from the DNA polymerase A catalytic subunit and the RNA polymerase II largest subunit. The phylogeny obtained provided considerable support for a Neotropical origin of the genus* Leishmania *and leads us to reject the hypothesis for a Palaearctic origin.*

Key words: biogeography - *Endotrypanum* - *Leishmania (Sauroleishmania)* - DNA polymerase - RNA polymerase

Contradictory hypotheses for either a Neotropical or a Palaearctic origin of the genus *Leishmania* have been proposed (Noyes et al. 1997, Kerr 2000). Kerr proposes that the mammalian *Leishmania* evolved from lizard *L. (Sauroleishmania)* in the Palaearctic which then migrated to the Nearctic in the Oligocene and to the Neotropics after the Isthmus of Panama had formed in the Pliocene (3-5mya). Noyes et al. (1997) and Noyes (1998) propose that the *Leishmania/Endotrypanum* clade evolved in the Neotropics during the first half of the Cenozoic (65-40mya), descendants of these parasites migrated through the Nearctic to the Palaearctic no later than the mid-Miocene. The lizard *L. (Sauroleishmania)* subsequently evolved from mammalian parasites in the Palaearctic.

TESTING BIOGEOGRAPHIC HYPOTHESES WITH INDEPENDENT EVIDENCE

How can these two contradictory hypotheses be tested? When studying biogeography, the number of scenarios that can be constructed to explain the data is limited solely by the imagination of the investigator. Consequently, the important (i.e. scientific) process is to explicitly test the scenarios with evidence that is independent of the construction of the biogeographic hypothesis in the first place. Historical biogeography involves the testing of explicit biogeographic hypotheses with quantitative evidence from the evolutionary history of the organisms (Myers & Giller 1988). That is, we expect considerable concordance between the geographic history and the phylogeny of the organisms under study. So, the congruence between the biogeographic patterns and the phylogenetic patterns is assessed, and the degree of congruence is used as a measure of the degree of support that the phylogeny provides for the biogeographic hypothesis.

From this perspective, Kerr (2000) provides no explicit test of the proposed biogeographic hypothesis, because no independent phylogenetic analyses are presented. This is particularly problematic, as the proposed phylogeny "is based on biogeography, the fossil records of mammals and sand flies, and ecology" — that is, the phylogeny is derived from the biogeographic hypothesis and therefore, should not be used as an independent test of that hypothesis (Malhotra et al. 1996).

PHYLOGENETIC ANALYSES USING A COMBINED SEQUENCE DATASET FROM DNA AND RNA POLYMERASE GENES

An appropriate test of the biogeographical hypothesis thus requires an independent estimate of the phylogeny of *Leishmania*. Such an estimate was provided for a number of *Leishmania* by Croan et al. (1997), for example, based on the nucleotide sequence of the DNA polymerase A catalytic subunit. Their phylogenetic tree is unrooted, although they provided several reasons for rooting the tree in the Neotropical clades so that *Endotrypanum* and *L. (Viannia)* are sister groups to the *L. (Leishmania)* species (Kerr incorrectly refers to these

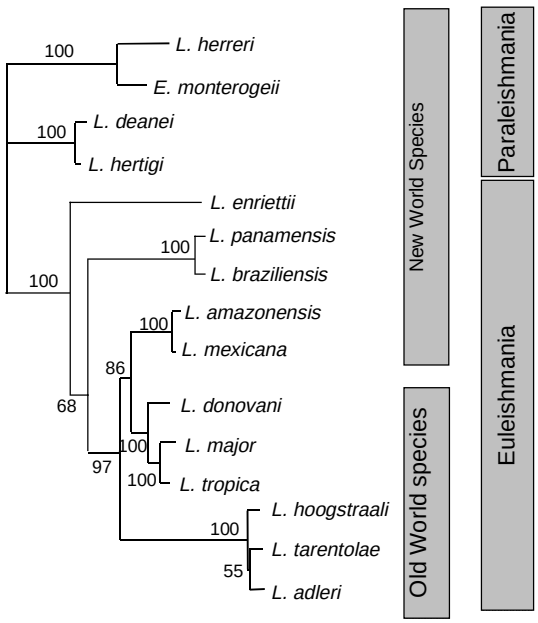
+Corresponding author. Fax: +61-2-95144003. E-mail: j.ellis@uts.edu.au
Received 13 April 2000
Accepted 15 May 2000

groups as "ancestral", whereas a cladogram shows only sister-group relationships rather than ancestor-descendant relationships). Thus, the independent phylogenetic data of Croan et al. (1997) provide no support for a Palaearctic origin of *Leishmania*. Kerr (2000) correctly states that the tree could "conceivably" be rooted "with *Sauroleishmania* and *L. (Leishmania)* at the base and *L. (Viannia)* and *Endotrypanum* at the crown". Although the phylogeny of Croan et al. (1997) was not rooted there are good reasons for placing its root amongst the Neotropical taxa (see below) and a number of rooted phylogenies of *Leishmania* have been published which show the Neotropical species at the root. For example, the rooted phylogenetic tree of Medina-Acosta et al. (1993), based on the nucleotide sequence of the surface proteinase (gp 63), has the Neotropical species at the root of the tree and the Palaearctic species at the crown. Similarly, the phylogenetic tree of Thomaz-Soccol et al. (1993), based on isoenzyme analysis of 13 different enzymes, is rooted with the Neotropical species at the base. This latter tree also cannot be rooted in any way that would support the phylogenetic tree required by Kerr's biogeographic hypothesis. Brewster and Barker (1999) present a rooted tree of the ATPase gene that also places the root of the *Leishmania* clade between *L. (Viannia)* clade and the *L. (Leishmania)/L. (Sauroleishmania)* clade.

We have extended the work of Croan et al. (1997), and present here a phylogenetic analysis of 2,171 nucleotides from protein-coding regions of 16 *Leishmania* and associated species. This phylogeny resulted from combining the RNA and DNA polymerase gene sequences into a single dataset for analysis and hence effectively doubles the amount of sequence data analysed, when compared to previous phylogenies. In the majority of cases adding data in this way has been shown to be more effective than adding taxa for resolving inconsistencies caused by unequal rates (Poe & Swofford 1999). Therefore this phylogeny of the *Leishmania* is, in principle, the most robust available to date.

The Figure shows the optimal tree obtained from maximum-likelihood analysis and this represents the current, best hypothesis of the historical relationships amongst the sequences analysed. This tree is perfectly compatible with the optimal tree shown in figure 1 of Croan et al. (1997). Since this optimal tree cannot be rooted to make the old world *Leishmania* monophyletic, this tree explicitly contradicts the phylogenetic hypothesis of Kerr (2000). Thus, the biogeographic hypothesis of Kerr (2000) fails this quantitative test, and must be considered refuted by these data. However, we note that inclusion of data from more species and related taxa in future analyses may result in one of the sub-optimal trees becoming the preferred maximum-likelihood tree, because there are a large number of sub-optimal trees that are not statistically significant different from the maximum-likelihood tree presented (Fig.).

Explicitly rooting the tree requires data from the sister group of the *Leishmania*. Potential sister species for the *Leishmania* include *Leptomonas* and *Crithidia*, with *Trypanosoma* being more distant (Briones et al. 1992, Maslov & Simpson 1995, Maslov et al. 1996). This is somewhat problematic, as only a small amount of data exists for these groups. The nearest relative for which complete data exist is *T. brucei*, but this cannot be aligned



Phylogenetic relationships among combined DNA polymerase A catalytic subunit (905 nucleotides) and RNA polymerase II largest subunit (1,266 nucleotides) sequences from selected species of *Leishmania* and *Endotrypanum*. The nucleotides were aligned using the Clustal algorithm in MEGALIGN (Saitou & Nei 1987, Higgins et al. 1989) in the DNASTAR package (DNASTAR, Inc. Madison WI, USA). An unrooted maximum-likelihood tree is shown, derived using the DNAML program of Felsenstein (1995), with the options for global rearrangements and five random starts. The branch lengths are drawn proportional to the evolutionary distances. Sequences are taken from Croan et al. (1997), except for the *L. enriettii* sequences [GenBank Accession Nos. AF151728 (DNA polymerase) and AF151727 (RNA polymerase)]. The numbers at the nodes show the bootstrap support for that node, based on 100 bootstrap replicates. The RNA and DNA polymerase phylogeny has the same topology as that of Croan et al. (1997); all the most basal species are Neotropical, whilst only derived species are found in the Old World. The position of *L. enriettii* suggests that cladogenesis was occurring within the genus *Leishmania ss* in the Neotropics before any of the Old World species had arisen.

reliably against the *Leishmania* data set. A 501-base-pair fragment of the RNA polymerase II largest subunit of *C. luciliae* aligns easily and thus allows a partial estimate of the root position (see Croan & Ellis 1996), indicating that the root should be on the branch leading to *L. enriettii*. Alternatively, using the assumption of a molecular clock, as suggested by Croan et al. (1997), the root is indicated as being on the branch connecting *Endotrypanum* + *L. herreri* + *L. deanei* + *L.hertigi* with the other species. Neither of these roots is compatible with a Palaearctic origin of *Leishmania.*

CHANGES IN EVOLUTIONARY RATE AMONGST *LEISHMANIA*

Changes in rate of evolution in different branches of the tree can cause distortion of the phylogeny. Kerr (2000) suggests that rate changes may have caused the observed tree to have arisen from her hypothetical tree. However this is highly unlikely since the distortions caused by rate changes are detectable by a range of tests and usually lead to a reduction in bootstrap support.

Kerr (2000) states that the polymerase phylogeny of Croan et al. (1997) was compiled assuming a molecular clock. This is incorrect; the phylogeny was compiled using DNAML, which does not assume a molecular clock, and then it was *tested* using the DNAMLK program, which does assume a molecular clock. Discrepancies between the phylogenies compiled under the different models are a good indicator of non-clock like behaviour. Only the *L. (Sauroleishmania)* appeared not to be evolving in a clock like manner, which is consistent with its long branch and relatively low bootstrap value. *L. (Sauroleishmania)* appeared to be evolving slightly faster than the other clades although the hypothesis of Kerr (2000) predicts that it should be evolving much more slowly than the other clades. The behaviour of all other taxa was consistent with a molecular clock, and hence with the position of the root suggested by Croan et al. (1997).

If the *L. (Sauroleishmania)* are evolving faster than the *L. (Leishmania)*, this could have caused the *L. (Sauroleishmania)* to have been drawn from within the *L. (Leishmania)* towards the outgroup by the long branch effect (Felsenstein 1988). The bootstrap value of 86 for the *L. (Leishmania)* clade is a consequence of the *L. (Sauroleishmania)* sometimes clustering within this clade. Alternative topologies in which the *L. (Sauroleishmania)* clusters with either the *L. donovani* clade or the *L. mexicana* clade are not rejectable (Croan et al. 1997). If *L. (Sauroleishmania)* clusters with the *L. donovani* clade, this would be consistent with an origin of the lizard parasites in the Old World from mammalian parasites and just a single migration between the Old and New Worlds. Other interpretations involving two migrations between the Palaearctic and Nearctic are also consistent with the polymerase phylogeny (Noyes 1998).

The other relatively low bootstrap value (68) in the polymerase gene phylogeny is for the *L. Viannia*/*L. (Leishmania)*/*L. (Sauroleishmania)* clade. This is caused by *L. enriettii* clustering with the *L. (Viannia)* clade in some bootstrap replicates, and it is not possible to group *L. enriettii* with confidence using this data. The position of *L. enriettii* may also have been distorted by long branch attractions.

The maximum likelihood test for clock like behaviour and the lowered bootstrap values caused by *L. enriettii* and *L. Sauroleishmania* demonstrate that changes in relative evolutionary rates are detected by standard phylogenetic methods. Croan et al. (1997) demonstrated that all other clades within their phylogeny were behaving in a clock like manner and all other bootstrap values in the phylogeny of Figure are at or close to 100%. These indicate that this phylogeny is robust and unlikely to have been distorted by the additional rate changes suggested by Kerr (2000).

BIOGEOGRAPHY

Kerr (2000) makes five specific objections to the hypothesis of Noyes et al. (1997) and Noyes (1998) for a Neotropical origin of the genus *Leishmania*. However Noyes et al. (1997) and Noyes (1998) did not make three of the propositions attributed to them, i.e. that *Leishmania* migrated to the Nearctic in porcupines; that porcupines migrated to the Palaearctic across the Bering land Bridge; or that the adaption of *L. hertigi* to porcupines was a mechanism of isolating the genus from *Endotrypanum.*

Kerr (2000) also states that the Neotropical origin hypothesis requires three independent adaptions to murid rodents, in the Neotropics, the Nearctic and the Palaearctic. The Neotropical origin probably requires a minimum of two adaptions, once in the Neotropics and once in the Nearctic. Given the ease with which the majority of species of *Leishmania* will infect murid rodents this does not appear to be a major evolutionary obstacle.

Lastly Kerr (2000) suggests the fossil evidence shows that the phlebotomine sand flies originated in the Palaearctic. However, the earliest fossil sand flies (120mya) were found in the Lebanon which was south of the Tethys sea during the Cretaceous and consequently was in Gondwana and not the Palaearctic (Lewis 1982). Before 120mya Phlebotominae had probably lived for a long time in Pangea from where separate sand fly faunas could have developed in the Neotropics and the Old World (Lewis 1982).

The absence of the GP46 gene from all *L. (Viannia)* tested and also from *L. enriettii* (Hanekamp & Langer 1991), is also adduced as evidence for a Palaearctic origin of *Leishmania*. However, the dendrograms used to support either a Palaearctic (Kerr 2000) or a Neotropical origin (Fig.) require that this character be lost the same number of times. This character therefore cannot be used as evidence for either a Neotropical or Palaearctic origin of *Leishmania*.

It is also stated that the Bering land bridge broke between the Oligocene and the Pliocene approximately 30-5mya. Although the Bering land bridge was occasionally inundated during this period it remained intact most of this time until the mid Pliocene (3.5mya) (Herman & Hopkins 1980).

CONCLUSIONS

The fundamental objection to the hypothesis of Kerr (2000) for a Palaearctic origin of the genus *Leishmania* is that no explicit phylogenetic test of the hypothesis is presented. Instead a dendrogram is constructed using data that was used to develop the hypothesis. There are also additional problems associated with the misinterpretation of the sand fly fossil data and the GP46 data, and the misrepresentation of the hypothesis of Noyes et al. (1997) and Noyes (1998).

Nevertheless, the hypothesis of Kerr (2000) was evaluated by phylogenetic analysis using a combined DNA sequence dataset derived from DNA and RNA polymerase genes. The phylogeny obtained provided considerable support for a Neotropical origin of the genus *Leishmania* and leads us to reject the hypothesis for a Palaearctic origin suggested by Kerr (2000). Noyes et al. (1997) and Noyes (1998) has speculated on when and how *Leishmania* might have evolved and how the observed modern distribution of the genus might have arisen. There is plenty of scope for further debate on these events, but such speculation will only make useful progress if it is firmly based on independent evidence.

REFERENCES

Brewster S, Barker DC 1999. The ATPase subunit 6 gene sequence predicts that RNA editing is conserved between lizard- and human-infecting *Leishmania*. *Gene 235:* 77-84.

Briones MRS, Nelson K, Beverley SM, Affonos HT, Camargo EP, Floeter-Winter LM 1992. *Leishmania tarentolae* taxonomic relatedness inferred from phylogenetic analysis of the small subunit ribosomal RNA gene. *Mol Biochem Parasitol 53*: 121-128.

Croan D, Ellis J 1996. Phylogenetic relationships between *Leishmania*, *Viannia* and *Sauroleishmania* inferred from comparison of a variable domain within the RNA polymerase II 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.

Felsenstein J 1995. PHYLIP (Phylogeny Inference Package) Version 3.572c, University of Washington, Seattle.

Felsenstein J 1988. Phylogenies from molecular sequences: inference and reliability. *Ann Rev Genet 22:* 521-565.

Hanekamp T, Langer PJ 1991. Molecular karyotype and chromosomal localization of genes encoding two major surface glycoproteins, gp63 and gp46/M2, hsp70, and beta-tubulin in cloned strains of several Leishmania species. *Mol Biochem Parasitol 48:* 27-37.

Herman Y, Hopkins DM 1980. Arctic oceanic climate in late Cenozoic time. *Science 209:* 557-562.

Higgins, Desmond G, PM Sharp 1989. Fast and sensitive multiple alignments on a microcomputer. *CABIOS 5*: 151-153.

Kerr SF 2000. Paleartic origin of *Leishmania. Mem Inst Oswaldo Cruz 95*: 75-80.

Lewis DJ 1982. A taxonomic review of the genus *Phlebotomus* (Diptera: Psychodidae). *Bull Br Mus Nat Hist (Ent) 45:* 121-209.

Malhotra A, Thorpe RS, Black H, Daltry JC, Wuster W 1996. Relating geographic patterns to phylogenetic processes. In PH Harvey, AJ Leigh Brown, J Maynard Smith & S Nee (eds), *New Uses for New Phylogenies*, Oxford University Press, Oxford, p. 187-202.

Maslov DA, Simpson L 1995. Evolution of parasitism in kinetoplastid protozoa. *Parasitol Today 11*: 30-32.

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.

Medina-Acosta E, Beverley SM, Russell DG 1993. Evolution and expression of the *Leishmania* surface proteinase (gp63) gene locus. *Infect Agents Dis 2*: 25-34.

Myers AA, Giller PS 1988. *Analytical Biogeography. An Integrated Approach to the Study of Animal and Plant Distributions*, Chapman & Hall, London, xiv + 578 pp.

Noyes, HA 1998. Implications of a Neotropical origin of the genus *Leishmania. Mem Inst Oswaldo Cruz 93:* 657-661.

Noyes HA, Arana BA, Chance ML, Maìngon 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.

Poe S, Swofford DL 1999. Taxon sampling revisited. *Nature 398:* 299-300.

Saitou N, Nei M 1987. The neighbor joining method: A new method for reconstructing phylogenetic trees. *Mol Biol Evol 4*: 406-425.

Thomaz-Soccol V, Lanotte G, Rioux JA, Pratlong F, Martini-Dumas A, Serres E 1993. Monophyletic origin of the genus *Leishmania* Ross, 1903. *Ann Parasitol Hum Comp 68*: 107-108.