Molecular trees of trypanosomes incongruent with fossil records of hosts

Sara F Kerr

Department of Biology, University of the Incarnate Word, 4301 Broadway, San Antonio, Texas 78209, US

Molecular trees of trypanosomes have confirmed conventionally accepted genera, but often produce topologies that are incongruent with knowledge of the evolution, systematics, and biogeography of hosts and vectors. These distorted topologies result largely from incorrect assumptions about molecular clocks. A host-based phylogenetic tree could serve as a broad outline against which the reasonability of molecular phylogenies could be evaluated. The host-based tree of trypanosomes presented here supports the “invertebrate first” hypothesis of trypanosome evolution, supports the monophyly of Trypanosomatidae, and indicates the digenetic lifestyle arose three times. An area cladogram of Leishmania supports origination in the Palaeartic during the Palaeocene.

Key words: Trypanosomatidae - evolution - Trypanosoma - Leishmania - Euleishmania - Paraleishmania

Trypanosomatidae (Euglenozoa: Kinetoplastida), which include the important human parasites Trypanosoma brucei, Trypanosoma cruzi, and Leishmania spp., also infect a broad array of invertebrates, vertebrates, and plants. Disagreement exists as to whether the first host was an invertebrate (Baker 1963) or a vertebrate (Wallace 1966), whether the family is monophyletic (Lukeš et al. 1997, Haag et al. 1998, Stevens et al. 1999, Hamilton et al. 2004) or paraphyletic (Fernandes et al. 1993, Maslov & Simpson 1995, Maslov et al. 1996, Hughes & Piontkivska 2003), the number of times the digenetic lifestyle has arisen (Hughes & Piontkivska 2003), and tree topology (Noyes 1998, Kerr 2000). Often molecular trees of trypanosomes are incongruent with knowledge of the evolution of hosts and vectors (Kerr 2000). These inconsistencies are due to inadequate representation of most genera (Hughes & Piontkivska 2003), inadequate rooting (Kerr 2000, Hughes & Piontkivska 2003), and incorrect assumptions about the molecular clock (Kerr 2000).

Many molecular phylogenies are inconsistent with the fossil records of hosts and vectors. For example, the placement of Trypanosoma vivax, a parasite of Artiodactyls, which originated 55 million years ago (Mya), basal to the outgroup Bodoniidae, a related kinetoplastid family containing free living parasites, monogenetic parasites of invertebrates, and digenetic parasites of leeches and fish (Hughes & Piontkivska 2003), suggests that all trypanosomes and bodonids are younger than 55 million years. The conclusion by Haag et al. (1998) that Salivaria diverged from other Trypanosoma up to 500 Mya is inconsistent with the Cenozoic origin of its placental mammal hosts and tsetse fly vectors. The placement of Sauroleishmania, a parasite of reptiles and sand flies, at the crown of a phylogenetic tree of Leishmania (Noyes 1998, 2000, Stevens et al. 2001) is completely incongruent with the information that reptiles (Carroll 1997) and phlebotomine sand flies (Lewis 1982) co-existed during the Cretaceous, prior to the evolution of placental mammals, which are currently hosts for Leishmania.

MATERIALS AND METHODS

A phylogenetic tree of trypanosomes and an area cladogram of Leishmania were constructed based on the evolutionary patterns and biogeography of hosts and vectors. The following assumptions were made: (1) general consensus exists on the classification of trypanosomes into the monogenetic genera Blastocrithidia, Crithidia, Herpetomonas, Leptomonas, Rhynchoidomonas, and Wallaceina; and the digenetic genera Endotrypanum, Leishmania, Phytomonas, and Trypanosoma (Baker 1963, Wallace 1966, Stevens et al. 2001, Grimaldi & Shottelius 2001, Hughes & Piontkivska 2003); (2) monogenetic parasites originated no earlier than current hosts; (3) digenetic parasites originated no earlier than the younger of its hosts; (4) dispersals of trypanosomes were congruent with dispersals of current hosts and vectors.

RESULTS

Host-based tree of trypanosomes - A host-based tree of trypanosomes is illustrated in Fig. 1. The tree contains two major branches: one containing Trypanosoma, which are parasites with vertebrate reservoirs and leech or insect vectors; the other containing monogenetic parasites of insects (Blastocrithidia, Crithidia, Herpetomonas, Leptomonas, Rhynchoidomonas, and Wallaceina) and the digenetic Endotrypanum, Leishmania, Phytomonas, and Sauroleishmania. Leeches and insects belong to the sister phyla Annelida and Arthropoda, both of which originated in the Cambrian (Boardman et al. 1987). There is no fossil record of leeches, but it is possible that free-living annelids that existed during the Cambrian were monogenetic hosts of trypanosomes. Hemiptera originated in the Pennsylvanian and Diptera in the Permian (Gullan & Cranston 2000), indicating possible origination times for their trypanosomes.

Financial support: US National Institutes of Health grant GM055337
E-mail: kerrmchugh@lstx.net
Received 19 July 2005
Accepted 4 January 2006
Leeches are common vectors of digenetic trypanosomes among fish and amphibians (Molyneux 1977); therefore, the first leech/vertebrate association probably goes back to the origin of fish in the Ordovician Period, followed by radiation with fish and amphibians. Digeneric trypanosomes of insects include Trypanosoma (insect/vertebrate), Sauroleishmania (Diptera/reptile), Leishmania (Diptera/mammal) Endotrypanum (Diptera/mammal), and Phytomonas (Hemiptera/plant) (Stevens et al. 2001). Camargo et al. (1990) reported that Leptomonas and Herpetomonas were also detected in plants; possibly they are also digenetic with insects. This branching pattern suggests that the digenetic life cycle in trypanosomes arose three times: (1) leech/fish; (2) Hemiptera/plant; and (3) Diptera/reptile. Alternatively, if Sauroleishmania is a direct descendant of a digenetic parasite acquired by Diptera from an infected plant, the digenetic life style arose only twice in trypanosomatids.

Area cladogram of Leishmania - A host-based area cladogram of Leishmania produces three possibilities, all showing the origin of Leishmania in the Palaeartic. The hypothesis of a Palaeartic origin is based on the fossil record of phlebotomine sand flies existing in the Palaeartic during the Cretaceous (Lewis 1982), the origin of placental mammals in this region (Cox & Moore 2000), and reasoning that the extreme isolation of the Neotropical throughout the Cenozoic offered little opportunity for dispersal elsewhere.

Fig. 2a is a modification of a phylogenetic tree presented by Kerr (2000) and Kerr et al. (2000) and illustrates the introduction of Leishmania into the Neotropical with murid rodents. Leishmania may have been introduced into the Neotropical with murid rodents during the Pliocene via the Panamanian land bridge, as was proposed by Kerr (2000) and Kerr et al. (2000) (Fig. 2a.), followed by a host switch to Caviomorphs (new world porcupines, guinea pigs, spiny rats and other South American rodents) and Xenarthra (sloths, armadillos and anteaters).

In Fig. 2b the antiquity of Leishmania in South America has been extended back from the Pliocene to the Oli-
Fig. 2: host-based area cladograms of *Leishmania*, which use patterns of origination and dispersal of hosts and vectors to infer the phylogeny of parasites. The following genera and subgenera are included: *Endotrypanum* (*Endo*); *Leishmania* (*Leishmania*) (*Leish* (*L.*)); *Leishmania* (*Viannia*) (*Leish* (*V.*)); unclassified *Leishmania* (*Leish* (?)); and *Sauroleishmania* (*Sauro*). Hosts include: Phlebotomine sand flies (Phleb); Reptiles (Rep); Murid rodents (Murid); Caviomorph rodents (Cavio); and Xenarthra (Xen). Sources for origination dates are mammals (Nowak 1991) and phlebotomines (Lewis 1982). *Leishmania* and *Endotrypanum* genera are based on Lainson and Shaw (1987), Lainson (1997), and Cupolillo et al. (2000). Alternative hypotheses for introduction of *Leishmania* into the Neotropical: (a) murid rodents in Pliocene; (b) hystricognath rodents in Oligocene; (c) ancestor of xenarthrans in Palaeocene.
gocene, in recognition that *Leishmania* may have been introduced into the Neotropical with hystricognath rodents prior to its introduction with murid rodents. It could have arrived with hystricognath rodents (Old and New World porcupines) via Africa, North America or Antarctica (Huchon & Douzery 2001), followed by radiation with caviomorphs and a host switch to Xenarthra (Fig. 2b). Host switches to xenarthrans may have then led to the evolution and radiation of *L.* (*Viannia*) and *Endotrypanum*. Possibly *L.* (*L.*) *mexicana* was introduced later when its murid rodent hosts arrived in the Pliocene.

Fig. 2c illustrates the possibility of the introduction of *Leishmania* into the Neotropical with the ancestor of xenarthrans. If *Leishmania* first arrived in the Neotropical with the ancestors of xenarthrans (Fig. 2c), it might have then radiated first with Xenarthrans and later with caviomorphs into *L.* (*Viannia*), Sauralezaishmania and *Endotrypanum*.

**DISCUSSION**

Comparison with molecular trees - Comparison of the host-based phylogenetic tree in Fig. 1 with a molecular tree using 18S rRNA, and which included free living, monogenetic, and digenetic Trypanosomatidae and Bodoniidae from a broad array of hosts and vectors (Hughes & Piontkivska 2003), shows the consistencies and the inconsistencies of the two methods. Both approaches show clusters containing the following groups: (1) monogenetic trypanosomes and the digenetic *Phytomonas, Leishmania*, and *Endotrypanum*; (2) Stercoraria and other non-salivarian *Trypanosoma* spp.; and (3) Salivaria. The maximum likelihood (ML) tree is most consistent with a host-based phylogeny, differing only by placing Salivaria external to other *Trypanosoma*. Minimum evolution (ME), maximum parsimony (MP), and Baysian trees also indicate that Salivaria is an early diverging group (Hughes & Piontkivska 2003), in contrast to the host-based tree which shows it evolved recently. *T. vivax* clustered outside all other Trypanosomatidae and Bodoniidae in ME and MP trees (Hughes & Piontkivska 2003), which greatly contrasts with the host-based tree by placing *T. vivax* ancestral to free-living species that could have evolved over 500 Mya. The ME, MP, and Baysian trees have highly asymmetrical bases, which is often an artifact of long branch attraction (Phillipe 2000). Since the ML tree corrects for long branch attraction, it probably best reflects the true topology. Hughes and Piontkivska (2003) concluded that trypanosomes are paraphyletic, in contrast to the host-based tree which indicates they are monophyletic. The conclusion that the transition to digenetic trypanosomes occurred multiple times is consistent with two or three transitions in the host-based phylogeny.

When compared to ML trees of SSU rRNA and gGAPDH genes (Hamilton et al. 2004) the host-based phylogeny of trypanosomes is consistent with both taxonomic groupings and topology of the SSU rRNA tree, but inconsistent with the gGAPDH tree. Hamilton et al. (2004) prefer the latter tree because it is based on more slowly evolving genes; however, the faster evolving SSU rRNA genes might reveal important details about fast evolving taxa. Their conclusion that trypanosomes are monophyletic and the implication that the digenetic life cycle arose more than once are consistent with the host-based tree, but their conclusion that an insect was the original host is inconsistent.

The inconsistencies of molecular phylogenies with host-based phylogenies are due to an overemphasis on the molecular clock in determining tree topology. Stevens and Rambaut (2001) recognized that there are differences in evolutionary rates between clades of trypanosomes. Vrba (1992) demonstrated that evolutionary rates also vary temporally and geographically, with accelerated rates of speciation occurring during times of climatic change, and in areas of greater topographic relief that provide more opportunities for allopatric speciation, such as the Neotropical. She concluded that when accelerated speciation events do occur, they impact many lineages simultaneously. From this it follows that the rates of evolutionary change of trypanosomes vary temporally and geographically, both within and between clades.

Seemingly, the salivarian mode of transmission by Diptera is a derived trait that evolved in parallel in *Trypanosoma* and *Sauroleishmania*, as adaptations to flies. Adaptation to the high body temperatures and complex immune systems of mammals likely resulted in additional distinctive patterns of accelerated evolution in the *T. cruzi, T. brucei*, and *Leishmania* clades. This accelerated rate of evolution causes these lineages to appear ancestral rather than derived, due to long branch attraction. The fact that ML trees (Hughes & Piontkivska 2003, Hamilton et al. 2004), which correct for long branch attraction, correspond best with the host-based tree (Fig. 1) suggests that these are the correct topologies.

*Murid, hystricognath or xenarthran introduction of Leishmania into the Neotropical?* - There are conflicting opinions about the origination and pattern of dispersal of *Leishmania* (Noyes 1998, 2000, Momen & Cupollo 2000, Stevens et al. 2001). The hypothesis of either a murid or hystricognath rodent as the transport host bringing *Leishmania* to the Neotropical is consistent with a host based phylogeny and with the discovery of a Cretaceous fossil record of *Leishmania* associated with reptiles by Poinar and Poinar (2004). *L. (L.) enriettii*, *L. (Viannia)*, and *Paraleishmania* all lack the GP46/M-2 gene family (Hanecamp & Langer 1991, McMahon et al. 1992, Cupollo et al. 2000); the most parsimonious explanation is a deletion occurring after the entry of *L. (Leishmania)* into the Neotropical in the Pliocene. If murid rodents first carried *Leishmania* to the Neotropical (Fig. 2a), the deletion occurred after *Leishmania* was introduced into the Neotropical with murid rodents in the Pliocene. Alternatively, the deletion could have occurred in the Oligocene after *Leishmania* was introduced with hystricognath rodents, or in the Paleocene after introduction with the ancestors of xenarthrans. In all cases, *L. (L.) enriettii* seems to be closest to the ancestral South American *Leishmania*.

The murid hypothesis (Fig. 2a) does not allow much time for radiation of South American *Leishmania*; in all other ways it is a very parsimonious explanation of the distribution of extant *Leishmania*. The close phylogenetic
The relationship of \( L. (L.) \) enriettii and Nearctic \textit{Leishmania} (ie \( L. (L.) \) mexicana) (Thomaz-Soccol et al. 1993) supports this hypothesis. The hystricognath (Fig. 2b) and xenartran (Fig. 2c) hypotheses both require two introductions of \textit{Leishmania} into the Neotropical, once with either hystricognath rodents or the ancestor of xenarthrans, and later with murid rodents. If hypothesis two is most accurate, then it might be expected that Old World porcupines would also be important reservoirs in the Old World, which they are not.

\textit{Natural selection of Leishmania} - Specific abiotic and biotic conditions are necessary to support \textit{Leishmania}, including optimum temperature and moisture, and the coexistence of both sand fly and mammal hosts, each with particular abiotic and biotic requirements. Intraspecific and interspecific interactions among \textit{Leishmania}, sand flies, and mammals add further complexity. One strong selective pressure on \textit{Leishmania}, in both sand flies and mammals, is temperature. In vitro, promastigotes of New World species are least tolerant of high temperatures and Old World species are most tolerant (Callahan et al. 1996). Limits of temperature tolerance of promastigotes are a good indicator of the limits of temperature tolerance of their sand fly hosts. Adaptation to hot climates that evolved in trypanosomes of insects or reptiles might have pre-adapted them to transition to the high temperatures of endothermic secondary hosts. Desert insects tolerate temperatures above 50°C and desert lizards tolerate temperatures up to 47°C, although temperatures during normal activity are much lower (Willmer et al. 2000).

Although mammals are probably secondary hosts of \textit{Leishmania}, they house the parasites much longer than the sand flies, since they may live many years, while survival of the adult sand fly, the only stage in which the parasite lives, may be only a few weeks. Furthermore, \textit{Leishmania} may persist in their vertebrate hosts during years when sand flies are absent due to adverse weather such as droughts. Therefore, adaptation to the vertebrate hosts must have contributed greatly to evolution of \textit{Leishmania}.

In vitro studies of the limits of temperature tolerance of \textit{Leishmania} amastigotes show that the upper limits of tolerance range from a low of 28°C in \textit{L. (V.)} bresiliensis (Eperon & McMahon-Pratt 1989) to a high of 39°C in \textit{L. (V.)} donovani (Zilberstein & Shapira 1994). Not unexpectedly, the limits of temperature tolerance of amastigotes corresponds with the body temperatures of their vertebrate hosts, reviewed by Willmer et al. (2000): 33-34°C in xenarthrans, sometimes dipping to 28°C in sloths (Gilmore et al. 2000); 36°C in marsupials; 37-38°C in rodents; and 38-39°C in primates, carnivores, and artiodactyls. Absence of avian \textit{Leishmania} may be due to the fact that the normal body temperature of birds exceeds 40°C (Willmer et al. 2000). Climate change throughout the Cenozoic, and the increase in body temperature throughout the evolution of mammals, likely resulted in a faster rate of evolution during this time period that distorts tree topology and gives the false impression that \textit{Sauroleishmania} is at the crown of the tree.

While molecular trees are extremely informative in revealing clusters of related species, the topologies of some molecular trees are nonsensical, with parasites of species that evolved recently basal to parasites of ancient species. This is most likely due to long branch attraction, with fast evolving species attracted to the base of the tree. There was a cooling trend in global temperatures in the Cenozoic culminating in the Quaternary Ice age of the last two million years (Cox & Moore 2000), resulting in rapid evolutionary change in recently evolved organisms that distorts the topology of molecular trees that assume a molecular clock with a constant rate of molecular change. For this reason, host-based phylogenetic trees may serve to check the reasonability of molecular trees.

The host-based tree of trypanosomes: (1) supports the “invertebrate first” hypothesis of trypanosomese evolution; (2) indicates the digenetic lifestyle arose two or three times, and that vertebrate hosts may have served as amplifying hosts, making the parasites widely available to a diversity of vectors, and therefore spreading the digenetic life cycle; (3) supports the monophyly of \textit{Trypanosoma}. The better correspondence of ML trees with host-based trees than other methods of molecular phylogenetic analysis affirms the importance of correcting for long branch attraction, and the superiority of ML trees over those that do not correct for it. An area cladogram of \textit{Leishmania}: (1) supports its origination in the Palaearctic during the Palaeocene; (2) supports introduction into the Neotropical with murid rodents in the Pliocene; and (3) indicates that the \textit{L. (L.)} enriettii lineage is basal to \textit{L. (Viannia), Paraleishmania, and Endotrypanum}. The evolution of trypanosomes likely extended back to the Cambrian, mirroring the evolution of eukaryotes: origination in the sea; transition to land; and diversification of trypanosomes along with their vectors and hosts due to the breakup of Pangaea and climatic fluctuations.

ACKNOWLEDGEMENTS

To CP McHugh for reviewing the manuscript.

REFERENCES


