The Evolutionary History of Kinetoplastids and Their Kinetoplasts

Alastair G. B. Simpson, Julius Lukeš,† and Andrew J. Roger**

*Department of Biochemistry, Canadian Institute for Advanced Research, Program in Evolutionary Biology, Dalhousie University, Halifax; and †Institute of Parasitology, Czech Academy of Sciences, and Faculty of Biology, University of South Bohemia, České Budějovice, Czech Republic

Despite extensive phylogenetic analysis of small subunit ribosomal RNA (SSUrRNA) genes, the deep-level relationships among kinetoplastids remain poorly understood, limiting our grasp of their evolutionary history, especially the origins of their bizarre mitochondrial genome organizations. In this study we examine the SSUrRNA data in the light of a new marker—cytoplasmic heat shock protein 90 (hsp90) sequences. Our phylogenetic analyses divide kinetoplastids into four main clades. Clades 1–3 include the various bodonid kinetoplastids. Trypanosomatids comprise the fourth clade. SSUrRNA analyses give vastly different and poorly supported positions for the root of the kinetoplastid tree, depending on the out-group and analysis method. This is probably due to the extraordinary length of the branch between kinetoplastids and any out-group. In contrast, almost all hsp90 analyses place the root between clade 1 (including *Dimastigella*, *Rhynchomonas*, several *Bodo* spp., and probably *Rhynchobodo*) and all other kinetoplastids. Maximum likelihood and maximum likelihood distance analyses of hsp90 protein and second codon–position nucleotides place trypanosomatids adjacent to *Bodo saltans* and *Bodo* cf. *uncinatus* (clade 3), as (weakly) do SSUrRNA analyses. *Hsp90* first codon– plus second codon–position nucleotide analyses return a slightly different topology. We show that this may be an artifact caused, in part, by the different evolutionary behavior of first- and second-codon positions. This study provides the most robust evidence to date that trypanosomatids are descended from within bodonids and that *B. saltans* is a close relative of trypanosomatids. A total reevaluation of the high-level systematics within kinetoplastids is needed. We confirm that the interlocking network organization of kinetoplast DNA seen in trypanosomatids is a derived condition within kinetoplastids but suggest that open-conformation minicircles may have arisen early in kinetoplastid evolution. Further understanding of the evolution of kinetoplast structure and RNA editing is hampered by a paucity of data from basal (i.e., clade 1) bodonids.

Introduction

Medically, ecologically, and evolutionarily speaking, Kinetoplastida (Eukaryota, Excavata, Euglenozoa) is one of the most interesting groups of eukaryotic microorganisms. Kinetoplastids include the trypanosomatid parasites responsible for major human maladies such as sleeping sickness (*Trypanosoma brucei*), Chagas' disease (*Trypanosoma cruzi*), and leishmaniases. Other kinetoplastids are collectively referred to as bodonids. Among these, the best studied are the cryptobiids (assigned, controversially, to *Cryptobia* and *Trypanoplasma*), some of which parasitize commercially important fish (Woo 2001). Free-living bodonids such as *Bodo* spp. doubtless play major roles in microbial food webs, consuming bacteria and small eukaryotes (Arndt et al. 2000).

Extensive research on trypanosomatids has revealed many unusual cellular, biochemical, and genomic traits (see Donelson, Gardner, and El-Sayed 1999). Trypanosomatid nuclear genomes appear remarkably streamlined for eukaryotes, with a high gene density, polycistronic transcription, overlapping genes, and almost no spliceosomal introns (Myler et al. 1999; Liniger et al. 2001). In contrast, their mitochondrial genome organization is baroque. The mitochondrial DNA is condensed into a massive body called the kinetoplast. Trypanosomatid kinetoplast DNA (kDNA) is a network of interlocked circular molecules of two classes, maxicircles and minicircles. The few maxicircles are similar in sequence and encode classical mitochondrial genes. The minicircles, numbering in the thousands, are open-conformation circles (not supercoiled) and are each interlocked with around three neighbors to form the continuous network (for reviews see Klingbeil et al. 2001; Lukeš et al. 2002). Depending on the taxon, each minicircle encodes 1–4 small guide RNAs (gRNAs) with dozens of different sequence types per cell. These gRNAs mediate drastic editing of many maxicircle gene transcripts by insertion and deletion of uridine residues. Editing usually occurs either at regions along the whole transcript (pan-editing) or at the 5′ end only (Estévez and Simpson 1999).

Attempts to understand the evolutionary history of trypanosomatids, especially the interplay between the acquisition of parasitism, variations in kDNA structure, and degree of RNA editing, have inspired examinations of kinetoplast organization in bodonids (Simpson and Maslov 1999; Simpson et al. 2000). Like trypanosomatids, many bodonids have a single kDNA mass, a eukinetoplast (see Vickerman 1977). In others, kDNA is distributed through the mitochondrion, either as one diffuse entity (pankinetoplast) or as distinct nodules (polykinetoplast). More detailed examinations reveal further diversity. *Bodo saltans* (free-living, with eukinetoplast) has open-conformation minicircles, each encoding two gRNAs (Blom et al. 1998, 2000). *Dimastigella* (essentially free-living, polykinetoplastidic) also has open-conformation minicircles (Štolba, Jirků, and Lukeš

Abbreviations: MLdistance, maximum likelihood distance; TBR, tree bisection–reconnection.

Key words: kinetoplast, parasites, phylogeny, protists, RNA editing, trypanosomes.

Address for correspondence and reprints: Alastair G. B. Simpson, Department of Biochemistry, Program in Evolutionary Biology, Canadian Institute for Advanced Research, Dalhousie University, Halifax, Nova Scotia, Canada B3H 4H7.
E-mail: simpson@hades.biochem.dal.ca.

Mol. Biol. Evol. 19(12):2071–2083. 2002

2001). *Cryptobia helicis* (commensal, pankinetoplastidic) and *Bodo caudatus* (free-living, eukinetoplastidic) reportedly have supercoiled minicircles (Hajduk, Siqueira, and Vickerman 1986; Lukeš et al. 1998). In *Trypanoplasma borreli* (parasitic, eukinetoplastidic) gRNAs are located on huge megacircles (Yasuhira and Simpson 1996; Simpson 1997). The minicircles or megacircles do not form a continuous network in any of these taxa. Actual RNA editing has been confirmed for *B. saltans*, *Bodo* cf. *uncinatus* and *T. borreli* (Lukeš et al. 1994; Maslov and Simpson 1994; Blom et al. 1998). Minicircles and RNA editing have not been detected in other Euglenozoa studied to date (euglenids and diplonemids).

Placing these findings in an evolutionary context (e.g., Simpson et al. 2000) is hampered by a poor understanding of kinetoplastid phylogeny. Morphological approaches have been cursory and nonrigorous (Kivic and Walne 1984). Phylogenies of several marker genes, primarily small subunit ribosomal RNA (SSUrRNA), have emerged for over a decade but have usually drastically undersampled bodonids (e.g., Landweber and Gilbert 1994; Maslov et al. 1994; Blom et al. 1998). This has been a critical shortcoming, given the belief that trypanosomatids descended from within the bodonids. Sampling has been improved substantially by some recent SSUrRNA analyses (Atkins, Teske, and Anderson 2000; Doležel et al. 2000), but the resulting phylogenies raise more questions than they answer. The important bodonid taxa *Bodo* and *Cryptobia* appear polyphyletic. Some trees suggest that trypanosomatids are closely related to *B. saltans* (Simpson et al. 2000; Lukeš et al. 2002), but this is contradicted by other analyses (Doležel et al. 2000). Most importantly, no convincing root for the kinetoplastid tree has emerged (contrast Maslov, Yasuhira, and Simpson 1999; Simpson and Maslov 1999; Wright et al. 1999; Atkins, Teske, and Anderson 2000; Doležel et al. 2000; Simpson et al. 2000; Lukeš et al. 2002). Even a rooting between trypanosomatids and bodonids is possible, according to the most detailed study to date (Doležel et al. 2000). Our phylogenetic understanding remains inadequate for estimating the evolutionary history of kDNA structure and RNA editing. Almost every imaginable phylogenetic arrangement of the taxa for which we have data remains possible, and it is unclear whether these organisms are a good representation of the overall diversity of bodonids.

In this study we reexamine kinetoplastid phylogeny using two markers: (1) the available SSUrRNA data, and (2) the cytosolic form of heat shock protein 90 (hsp90). Cytosolic hsp90 is an essential molecular chaperone, seemingly universal in eukaryotes (Pearl and Prodromou 2000). Using degenerate PCR, we have determined near-complete sequences for several bodonid kinetoplastids and some out-groups. This study represents the first widespread survey of any nuclear-encoded protein gene in kinetoplastids and the first detailed phylogenetic examination with a taxon sampling approaching to that of SSUrRNA.

Materials and Methods

Gene Amplification, Cloning, and Sequencing

For most of the taxa examined (see below) genomic DNA was isolated as described previously (see Atkins, Teske, and Anderson 2000; Doležel et al. 2000). Throughout this study, the species identities used for examined strains was according to the above-mentioned authors. *Naegleria gruberi* ATCC 30224 DNA was a kind gift from R. Redfield (University of British Columbia). *Rhynchopus* sp. ATCC 50230 was grown axenically in seawater enriched with 10% w/v horse serum. Genomic DNA was extracted using a Puregene DNA isolation kit (Gentra).

Degenerate primers were designed against near-terminal conserved regions of known hsp90 protein sequences from other eukaryotes, including trypanosomatids and other Excavata (e.g., diplomonads). Primer design took into account biased codon usage (generally, there is an excess of C and G at degenerate positions in *hsp90* genes from Excavata). The forward primer 100XF was designed against QLMSLIINTFY (corresponding to residues 13–23 in *T. cruzi* hsp90) and had the 5′–3′ sequence CAGCTGATGTCCCTGATCATYA AYACNTTYTA. The reverse primer 970XR was designed against MIKLGLSLD/E (positions 666–675 in *T. cruzi*), with the sequence TCGAGGGAGAGRCCNA RCTTRATCAT. An alternate reverse primer, 968XR, was designed against IHRMIKLG (positions 663–671 in *T. cruzi*), with the sequence CCCAGCTTGATCAT NCGRTGDAT. An internal reverse primer, 430XR, was designed against WTRDPKDVT (positions 272–281 in *T. cruzi*), with the sequence GTGACCTCCTTGGGRT CNCGNGTCCA.

Near-complete (>90%) *hsp90* genes from *B. saltans* (Konstanz strain), *Bodo* cf. *uncinatus*, *Cryptobia salmositica*, *C. helicis*, *Dimastigella trypaniformis* (Ulm strain), *Diplonema papillatum*, *Rhynchomonas nasuta* (CBR1 strain), *Rhynchopus* sp. ATCC 50230, and *T. borreli* were amplified by PCR using 100XF and 970XR. Near-complete *hsp90* genes from *Bodo saliens* were amplified using 100XF and 968XR. A partial sequence was amplified from *N. gruberi* using 100XF and 430XR. A specific primer was then used with 970XR to obtain the rest of the gene (the sequences overlapped perfectly over 713 bases). Annealing temperatures were 48–55°C.

All amplification products were gel purified and cloned in pCR 2.1-TOPO TA cloning vector (Invitrogen, San Diego, Calif.). In most cases, two to five individual clones were partially sequenced, with limited (<1%), and mostly silent, heterogeneity detected between clones. A single clone was then selected for complete bidirectional sequencing. With *C. helicis* and *B. saliens* moderate sequence heterogeneity was detected among nine and seven clones, respectively. Two dissimilar clones from each were completely sequenced. All sequencing was performed using an ABI-377 automated sequencer.

Table 1
Summary of Phylogenetic Analyses Performed in This Study

SSUrRNA Data set	Sites (nucleotides)	Included Taxa
rK	1,521	Kinetoplastids
rKD	1,275	Kinetoplastids, diplonemids
rKDE	1,075	Kinetoplastids, diplonemids, euglenids
rKE	1,075	Kinetoplastids, euglenids
rKDES	1,020	Kinetoplastids, diplonemids, euglenids, short branches
rKDEH	1,020	Kinetoplastids, diplonemids, euglenids, Heterolobosea
rKDESH	1,020	Kinetoplastids, diplonemids, euglenids, Heterolobosea, short branches
Hsp90 Data set	**Sites (amino acids)**	**Included Taxa**
pK	605	Kinetoplastids
pKDHP	605	Kinetoplastids, diplonemids, Heterolobosea, plants
pKD	605	Kinetoplastids, diplonemids
pKDH	605	Kinetoplastids, diplonemids, Heterolobosea
pKDP	605	Kinetoplastids, diplonemids, plants
pKHP	605	Kinetoplastids, Heterolobosea, plants

Phylogenetic Analysis

SSUrRNA

SSUrRNA genes were aligned by eye. The taxon selection was similar to that used by Doležel et al. (2000) but with some newly available bodonid sequences added and some highly similar sequences omitted (accession numbers and alignments available at http://hades.biochem.dal.ca/Rogerlab). A series of out-group taxa were considered: other Euglenozoa (i.e., diplonemids and euglenids), Heterolobosea (*N. gruberi* and *Tetramitus rostratus*), and a small cluster of short-branching taxa (plant, stramenopile, alveolate, and *Trimastix*).

From this alignment we constructed several data sets with different taxon and site inclusions (see table 1). There were four primary data sets: (1) rK included kinetoplastids only and retained 1,521 unambiguously aligned sites, (2) rKD included diplonemids and kinetoplastids and 1,275 sites, (3) rKDE included all Euglenozoa and 1,075 sites, and (4) rKE retained the same 1,075 sites but included kinetoplastids and euglenids only. Three secondary data sets were constructed, all using 1,020 sites. These included (1) Euglenozoa plus Heterolobosea (rKDEH), (2) Euglenozoa plus short-branching taxa (rKDES), and (3) Euglenozoa plus both Heterolobosea and short-branching taxa (rKDESH).

All data sets were subjected to maximum likelihood analysis in PAUP* b.08 (Swofford 2000). Using a maximum parsimony tree, the parameter values were estimated for a general time reversible model of nucleotide substitution, with a six-category discrete approximation of a gamma distribution plus invariable sites (hereafter, a GTR+Γ+I model). In several cases the proportion of invariable sites was estimated to be zero, with constant or slowly evolving sites accommodated by a very low value for the α parameter. Empirical base frequencies were used. The best tree was searched for using 20 random taxon additions, with tree bisection–reconnection (TBR) rearrangements. For the primary analyses only, a maximum likelihood bootstrap analysis was also performed—1,000 replicates, each searched with three random taxon additions and TBR for the rK data set; 250 replicates, each searched by neighbour-joining followed by TBR for the other three data sets.

For both the primary and secondary data sets, maximum likelihood distance (MLdistance) analyses were also performed. These procedures were identical to those of the full maximum likelihood analysis, except that the gamma distributions were approximated by eight discrete categories; the search for the best tree used 100 random additions with TBR, and a 2,000 replicate bootstrap analysis was performed in all cases (five random additions with TBR per replicate).

SSUrRNA sequences from included kinetoplastids have similar G+C contents (48%–52% overall, 44%–52% for variable sites in rK). We saw no evidence of base composition artifacts with the rK data set; preliminary Logdet distance analyses and MLdistance analyses with the most G+C divergent taxa omitted returned topologies and support values very similar to those given by the full MLdistance analysis.

Hsp90

Conceptual amino acid translations of the *hsp90* gene sequences were aligned by eye. No introns were detected in any of the kinetoplastid or diplonemid sequences reported. One intron was manually spliced from the *N. gruberi* sequence. This intron inserted at phase 1 of codon 70 of the amplified region (corresponding to residue 93 of the complete *T. cruzi* sequence) and was 101 bases long. Four protein sequences from trypanosomatids were obtained from GenBank and were also aligned (*T. cruzi*, *T. brucei*, *Leishmania mexicana amazonensis*, *Leishmania infantum*). Two sequences from land plants were also included (earlier analyses have shown these to be short branches—data not shown). There were no publicly available hsp90 sequences for euglenids, and our attempts to amplify them were unsuccessful. Aside from the ends of the gene (the 100XF and 970XR priming sites and beyond) only two regions were ambiguously aligned and therefore omitted. These corresponded to residues 158–165 and 220–249 in *T. cruzi*. A total of 605 sites were retained for analysis.

The corresponding DNA sequence alignment was also assembled.

For the hsp90 data, a series of analyses were performed with different taxa included (table 1). Three primary data sets were examined. The first included only the kinetoplastids (pK). The second included diplonemids, Heterolobosea (*N. gruberi*), and the plants as outgroups in addition to kinetoplastids (pKDHP). The third included kinetoplasts and diplonemids only (pKD). Three secondary data sets were subjected to more cursory analysis. Each included only two of the three outgroup taxa (i.e., diplonemids and Heterolobosea, pKDH; diplonemids and plants, pKDP; Heterolobosea and plants, pKHP). The amino acid alignments showed little compositional heterogeneity, with all sequences easily passing chi-square heterogeneity tests in TREEPUZZLE 5.0 (Strimmer and von Haeseler 1996), even when only variable sites were considered.

For the three primary data sets, the amino acid alignments were analyzed both by maximum likelihood and by MLdistance. For maximum likelihood, TREEPUZZLE was used to estimate the parameter values for a six-category approximation of a gamma distribution plus invariable sites model of among-site rate variation. Empirical amino acid frequencies and the PAM substitution model were used (the PAM matrix is the only one supported by the software used for the maximum likelihood tree-searching described below). The proportion of invariable sites and the rates for the six discrete categories were inputted into PROML v3.6 (Felsenstein 2000), and the maximum likelihood tree was searched for using 10–20 random taxon addition sequences, each followed by global rearrangements. Bootstrap analyses were performed using the same settings but with one taxon addition sequence per replicate. Five-hundred replicates were performed for the pK and pKD data sets, whereas 250 replicates were performed for the pKDHP data set. For the MLdistance analysis, maximum likelihood estimates of pairwise distances were calculated in TREEPUZZLE using the JTT substitution matrix and an eight-category discrete approximation of a gamma distribution plus an invariable sites category. The best tree was found using FITCH 3.57 (Felsenstein 1995), with 100 random addition sequences and global rearrangements. Bootstrap analyses with 2,000 replicates were performed using PUZZLEBOOT (http://www.tree-puzzle.de/puzzleboot.sh) and FITCH (five additions with rearrangements per replicate).

The DNA sequences were analyzed by maximum likelihood using PAUP*, examining both first- and second-codon positions (1+2 analyses), and second-codon positions only (2-only analyses). Using a maximum parsimony tree, the parameter values were estimated for a GTR+Γ+I model, with an eight-category discrete approximation of a gamma distribution and empirical base frequencies. Tree searching used 10 random addition sequences, each followed by TBR. Bootstrap analyses were performed with 1,000 replicates using identical settings, except that there were only three random additions per replicate.

Similar analyses were performed on the secondary data sets, except that protein maximum likelihood bootstraps were not performed, and the DNA bootstrap analyses included only 250 replicates (starting trees obtained by neighbor-joining).

To examine the topological discrepancy between the maximum likelihood trees according to 1+2 analyses and trees derived with other data and methods (see *Results* below), we performed a series of likelihood calculations on two test trees and the three primary nucleotide data sets. Likelihoods were computed using the tree scores functions in PAUP* (see *Results*).

Results

SSUrRNA Data

Unrooted Analyses

Our analyses of the kinetoplastids-only data set (rK) return a fairly well-resolved tree (fig. 1*a*). In both maximum likelihood and MLdistance analyses, kinetoplasts fall into four well-supported clusters. We hereafter refer to these clusters as clades because our hsp90 analyses suggest that each cluster is indeed monophyletic (see below). Clade 1 contains *Bodo designis*, *B. saliens*, *Cruzella marina*, *Rhynchobodo* sp., *R. nasuta*, and *D. trypaniformis*. Clade 2 contains both *C. helicis* and the fish cryptobiids (*C. salmositica* and *T. borreli*), plus *B. caudatus*, *Procryptobia* (*Bodo*) *sorokini*, and *Parabodo nitrophilus*. Clade 3 contains both *B. saltans* isolates plus *Bodo* cf. *uncinatus*. The fourth clade corresponds to the trypanosomatids. Clades 1 and 3 and the trypanosomatids are all supported by bootstrap percentages of 97% or greater with both methods. Clade 2 is a little less robust (71% bootstrap support with likelihood, 85% with MLdistance).

The relationships within each clade are largely but not completely resolved. Within trypanosomatids there is a basal split between two well-supported clades, *Trypanosoma* on one hand and all other included trypanosomatids on the other. Within clade 1, *C. marina*, *B. designis*, and *B. saliens* form one strong clade, and *Rhynchomonas* and *Dimastigella* form a second clade. *Rhynchobodo* sp. is generally placed either as the sister to *Rhynchomonas* and *Dimastigella* (the maximum likelihood tree, as well as the majority of likelihood bootstrap replicates and 38% of distance bootstrap replicates) or as the basal branch within clade 1 (the best distance tree and 32% of distance bootstrap replicates). Within clade 2, the fish cryptobiids form a strong clade, as do *C. helicis*, *B. caudatus*, and *P. nitrophilus*. *Procryptobia sorokini* falls either as the sister of the fish cryptobiids (maximum likelihood) or as the sister of the *C. helicis*/*B. caudatus*/*P. nitrophilus* clade (MLdistance). Within clade 3, there is weak-to-moderate support for a clade of *B. saltans* isolates to the exclusion of *Bodo* cf. *uncinatus*. The SSUrRNA trees confirm that our taxon sampling with hsp90 is reasonably broad and well balanced (see fig. 1*a*).

The relationships among the four major clades are weakly resolved in the unrooted SSUrRNA analyses. Both analyses group clades 1 and 2 together and group

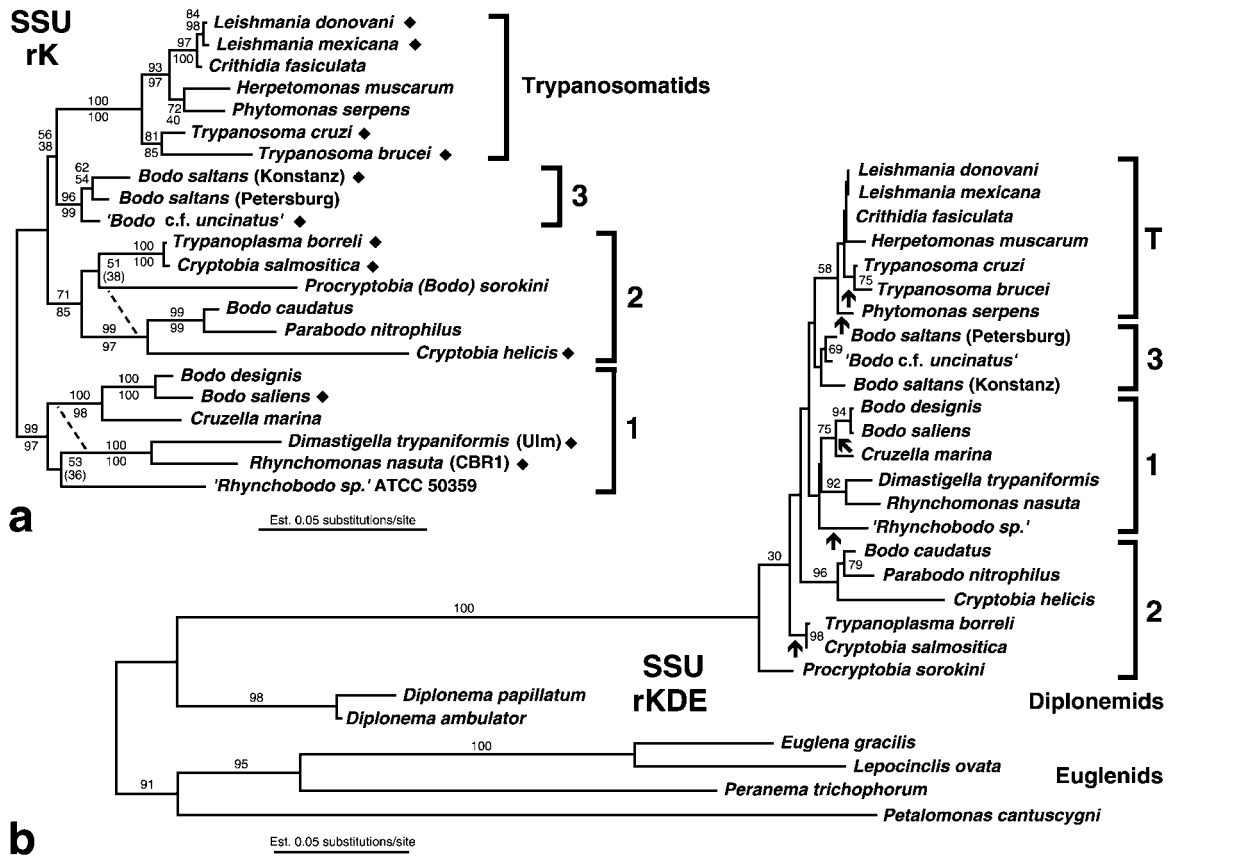

FIG. 1.—SSUrRNA phylogenies of kinetoplastids. (*a*) Maximum likelihood tree for kinetoplastids only (rK analysis), showing four well-supported clades. Tree is rooted in the position suggested by hsp90 phylogenies (see fig. 3). Dashed lines indicate topological elements that differ in the MLdistance analysis. Bootstrap support values are shown associated with internal branches (maximum likelihood top, MLdistance bottom). Numbers in parentheses signify a bipartition absent from the best MLdistance tree. Diamonds denote taxa for which hsp90 data is available (*Leishmania donovani* is a very close relative of *L. infantum*). (*b*) Maximum likelihood tree from the rKDE analysis, with other Euglenozoa used as out-groups. All bootstrap support values >50% are shown. The length of the branch leading to kinetoplastids and its low bootstrap support (30%) are typical for the analyses with out-groups. Arrows denote positions for the kinetoplastid root with other methods and out-group sets.

clade 3 with trypanosomatids, but the bootstrap support is weak (56% with likelihood, 38% with MLdistance; fig. 1*a*).

Rooted Analyses

We performed a series of further analyses to root the kinetoplastid tree by including sequences from other eukaryotes. Initially, we used diplonemids as the only out-group (analysis rKD), maximizing the number of alignable sites but artificially lengthening the branch between kinetoplastids and the out-group. To break this branch, we also added another out-group in euglenids, though this allows fewer sites to be included (analysis rKDE). We also tried to root kinetoplastids with euglenids alone (rKE). Previous SSUrRNA analyses often root Euglenozoa such that kinetoplastids are the sister to a clade of diplonemids and euglenids (see Moreira, Lopez-Garcia, and Rodriguez-Valera 2001). Therefore, the inclusion of out-groups to Euglenozoa might further break the basal branch of kinetoplastids. We tried two additional out-groups separately and in combination: Heterolobosea, a long-branching clade thought to be the sister group to Euglenozoa (e.g., Baldauf et al. 2000) and a small cluster of short-branching eukaryotes (analyses rKDEH, rKDES, and rKDESH).

The most striking feature of all the rooted analyses is that the basal branch leading from the out-groups to the kinetoplastids is extremely long (fig. 1*b*). It either is the longest internal branch in the tree or is a close second. It is always several times longer than any branch within kinetoplastids and ~20 times the length of the longest internal branch in kinetoplastids. In analyses where out-groups to Euglenozoa are included (rKDEH, rKDES, and rKDESH), diplonemids always form the sister group to kinetoplastids, with high MLdistance bootstrap support (70%–88%, maximum likelihood bootstraps not attempted). Thus, these out-groups to Euglenozoa did not, in fact, break the kinetoplastid branch (figure not shown).

The kinetoplastids are rooted in widely differing positions in different SSUrRNA analyses (see fig. 1*b*). *Rhynchobodo* sp., from clade 1, is basal in rKD distance analyses. In the rKD maximum likelihood tree the basal node is a trichotomy inside clade 1 between *B. designis* + *B. saliens*, *Cruzella*, and other kinetoplastids. The clade 2 *P. sorokini* is basal in the rKDE, rKDES, and

rKDESH likelihood analyses. The fish cryptobiids, also from clade 2, are basal in the rKDEH, rKDES, and rKDESH distance analyses. *Trypanosoma* is basal in the rKDEH likelihood analysis and the rKE distance analysis, whereas another trypanosomatid, *Phytomonas serpens*, is basal in the rKE likelihood analysis. The bootstrap values for these positions are always low (52% with rKD likelihood, otherwise 38% or less). Available SSUrRNA data seem ill suited to determining the position of the root of the kinetoplastid tree.

In all of our rooted trees, the relationships among kinetoplastids are more poorly resolved than those in the rK analysis. Ignoring the position of the kinetoplastid root (see below), the maximum likelihood trees generally retrieve the four clades seen previously, but one or more of these (especially 1 and 2) are nonmonophyletic in some MLdistance trees. Where tested, the bootstrap values supporting the four clades tend to be lower than those in the rK analysis with both methods (see fig. 1*b*).

Hsp90 Data

Unrooted Analyses

Our analysis of the kinetoplastids-only data set (pK) yields trees that are highly congruent to those from the rK SSUrRNA analyses (fig. 2). Both protein and 1+2 nucleotide analyses return the same four main clades as found with rK. Bootstrap support values are >85% for all clades, except for clade 3, which receives a 66% support value in the 1+2 nucleotide analysis. The 2-only nucleotide analysis gives similar results, except that clade 2 receives just 58% bootstrap support, whereas the maximum likelihood tree (but not the bootstrap consensus tree) actually places *B. saltans* and *Bodo* cf. *uncinatus* as successive branches at the base of trypanosomatids. Throughout this study, the 2-only analyses tend to give more weakly supported topologies than do other methods, as might be expected, given that it is the most information-poor data set examined.

The relationships within the clades are also largely in agreement with the rK analysis. Within clade 1, *Dimastigella* and *Rhynchomonas* usually group to the exclusion of *B. saliens*, although this node is only moderately well supported. Within clade 2, the fish cryptobiids always group strongly to the exclusion of *C. helicis*. Available taxon sampling does not allow us to confirm whether any free-living clade 2 organisms fall between the two groups of cryptobiids. Within trypanosomatids, *Trypanosoma* always forms a strong clade, as does *Leishmania*.

One interesting feature of the pK trees is the relationship among the four major clades. The protein and 2-only nucleotide analyses support the same internal partition as the SSUrRNA analyses, placing clade 1 adjacent to clade 2, and clade 3 with trypanosomatids (fig. 2*a*). Whereas this node is weakly supported in the SSUrRNA analysis, both the maximum likelihood and the MLdistance analyses of hsp90 proteins strongly support this partition (91% and 87% bootstrap support, respectively). The 2-only nucleotide analysis also provides moderate bootstrap support (73%). However, the maximum likelihood tree for 1+2-position nucleotides contradicts other analyses, placing clades 2 and 3 together (fig. 2*b*). Bootstrap support for this placement is weak with the pK data set (47%), but the same bipartition is returned by 1+2 nucleotide analyses when out-groups are included, often with higher support (70%+; see below and fig. 3). These anomalous results are explored further below.

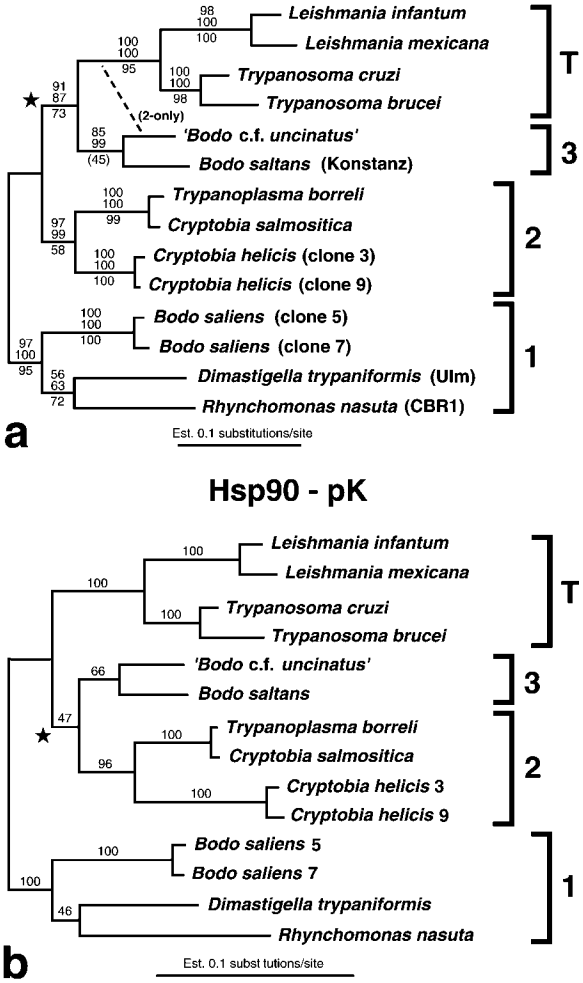

FIG. 2.—Hsp90 phylogenies of kinetoplastids with no out-groups considered (pK analysis). Trees are shown as rooted based on subsequent analyses (see fig. 3). The four well-supported major clades are denoted as in figure 1. (*a*) Maximum likelihood tree for proteins. Bootstrap support values are for maximum likelihood with proteins, MLdistances with proteins, and maximum likelihood with second-position nucleotides (2-only). Note the relatively strong support for the placement of clade 3 with trypanosomatids (asterisk). Dashed line indicates a topological element differing in the best tree from the 2-only analysis. (*b*) Maximum likelihood tree for first and second-position nucleotide data (1+2), with bootstrap support values shown. Note the position of clade 3 with clade 2, rather than with trypanosomatids (asterisk).

Rooted Analyses

We performed several analyses to root the kinetoplastid tree. Initially, we used diplonemids, Heterolobosea, and plants as out-groups in the pKDHP analysis (fig. 3*a*). Irrespective of the method, kinetoplastids form

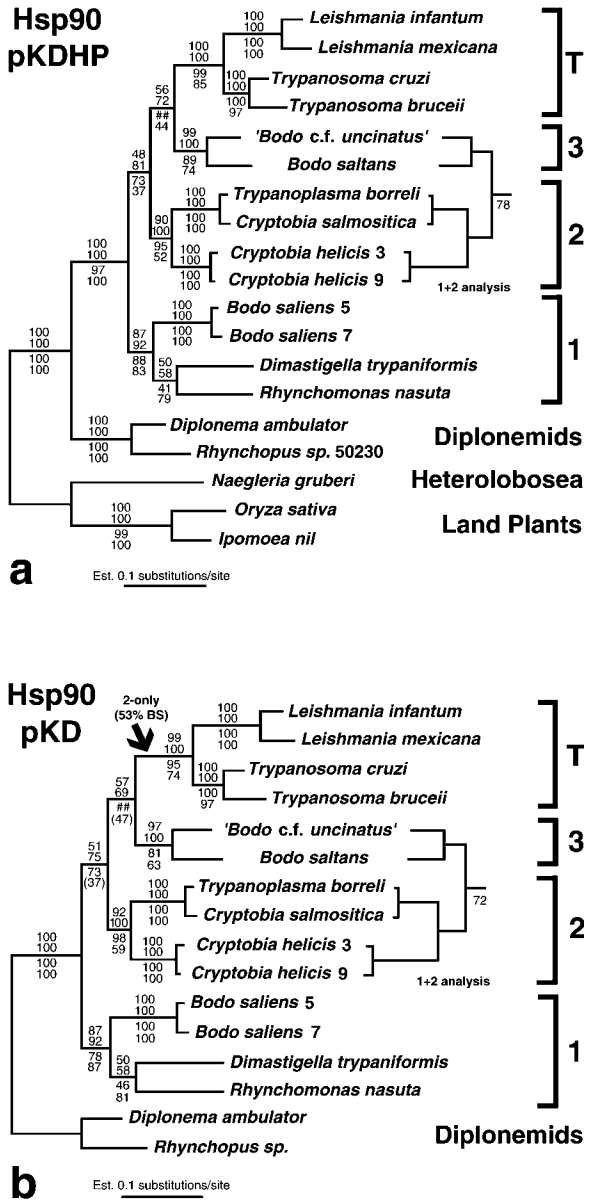

FIG. 3.—Hsp90 phylogenies of kinetoplastids and out-groups. Bootstrap support values are for protein maximum likelihood, protein MLdistances, maximum likelihood with 1+2 nucleotides and maximum likelihood with 2-only nucleotides. Numbers in parentheses signify substantial support for an element not in the best tree for an analysis. ## indicates an element absent from the best tree for an analysis and attracting little bootstrap support. (*a*) Maximum likelihood tree for protein sequences for the pKDHP data set. The topological element where the 1+2 nucleotide analysis disagrees with other methods is shown in the reflected subtree (with the bootstrap support for this element at its base). (*b*) Maximum likelihood tree for protein sequences for the pKD data set. Reflected subtree indicates the best 1+2 nucleotide topology as before. Arrow shows the position of the kinetoplastid root in the best tree from the 2-only analysis.

a strongly supported clade. Diplonemids always form the sister group to kinetoplastids, again with strong support. In contrast to SSUrRNA trees, the branch separating diplonemids and kinetoplastids is of a similar length to the internal branches within kinetoplastids. The longest interior branch is that separating diplonemids and kinetoplastids from *N. gruberi* and plants. *Naegleria* is always the longest terminal branch, especially with 1+2 nucleotides. All methods give the same unrooted topologies within kinetoplastids as with the pK data set, except that the 2-only analysis now recovers clade 3 as monophyletic. This stability of the unrooted topology of kinetoplastids is a constant in the analyses described below.

In contrast to the SSUrRNA analysis, all methods placed the root for kinetoplastids in the same place, between clade 1 and the rest of the kinetoplastids. This placement is strongly supported by the protein MLdistance and the 1+2 nucleotide analyses (81% and 75% bootstrap support, respectively). Bootstrap support is weaker with protein maximum likelihood and 2-only nucleotides (48% and 37% respectively).

To examine whether the rooting of the kinetoplastids is sensitive to the out-group taxa selected, we performed a series of analyses where one of the three clusters of out-groups was eliminated, leaving either diplonemids and plants only (pKDP), diplonemids and Heterolobosea only (pKDH), or plants and Heterolobosea only (pKHP). Again, in contrast to the SSUrRNA analyses, these analyses almost invariably return the same rooting position for kinetoplastids as when all taxa were considered. Protein MLdistance analyses consistently give high bootstrap support for this position (79%–82%). Due to computational constraints we did not do protein likelihood bootstrap analyses for these data sets. Bootstrap support with 1+2 nucleotides is ~66% in the pKDP and pKDH analyses. In the pKDP analysis, the bootstrap support is 44%, but the support values tended to be low across the tree. Perhaps the long *Naegleria* branch, not broken by the plants here, has a general destabilizing effect (figure not shown). With 2-only nucleotides, the pKDH and pKHP analyses give the same rooting as do other analyses (42% and 52% bootstrap support, respectively). The best 2-only nucleotide tree from the pKDP analysis actually roots kinetoplastids between trypanosomatids and bodonids, with very weak bootstrap support (22%). A rooting between clade 1 and other kinetoplastids actually garners more support (28%).

Finally, in the pKD analysis, diplonemids are used as the only out-group, based on their apparent closer proximity to kinetoplastids (fig. 3*b*). This use of a single out-group artificially lengthens the basal kinetoplastid branch but excludes the relatively long branches associated with the plants and *Naegleria*. Protein likelihood, protein MLdistance, and 1+2 nucleotide analyses, all root the kinetoplastids between clade 1 and the other taxa (fig. 3*b*). Bootstrap support values are rather similar to those from the pKDHP analysis (51%, 75%, 73%, respectively). The 2-only nucleotide analysis again roots kinetoplastids between trypanosomatids and bodonids. Bootstrap support is 53%, with much of the remaining support (37%) going to a rooting between clade 1 and other kinetoplastids (fig. 3*b*).

Artifacts in the First- Plus Second-Position Nucleotide Analyses?

We are interested in understanding why the 1+2 nucleotide analyses reconstruct the relationships among

Table 2
Estimated Relative Rates of all Reversible DNA Substitutions for First- and Second-Nucleotide Positions from *hsp90* Genes

	A↔C	A↔G	A↔T	C↔G	C↔T	G↔T
1 + 2 positions . . .	1.33	0.87	0.69	0.98	1.26	0.86
Second positions . .	0.97	1.18	0.26	2.76	0.61	0.21
First positions.	1.11	0.55	1.23	0.28	1.65	1.18

NOTE.—The first row shows the estimated rates when both positions are considered together. The last two rows show the estimated rates when first and second positions are considered separately. The rates shown are estimated according to the 1+2nuc tree using the pK dataset, with all parameters reestimated (i.e., the first row in Table 3). Similar values were obtained in all analyses, including the original maximum likelihood tree searching and bootstrapping.

the four main clades of kinetoplastids differently to the protein and 2-only nucleotide analyses. There are large differences in the DNA substitution matrices estimated for 1+2 nucleotides and 2-only nucleotides, with the relative rate for a substitution being up to ~4 times higher in one matrix than in the other (see table 2). This suggests that the matrices used for the 1+2-position analyses are poor approximations of the substitution processes occurring both at second positions and at first positions. Could it be that this dubious substitution model is causing an error in the phylogenetic inference?

Using PAUP*, we used a simple approach to examine whether allowing different models of sequence evolution at first- and second-nucleotide positions affected the relative likelihoods of different trees (see Wilgenbusch and de Queiroz 2000). We used our three primary hsp90 data sets (pK, pKD, and pKDHP) and examined two different trees: the maximum likelihood tree for the 1+2 nucleotide data, which places clade 3 with clade 2 (hereafter the 1+2nuc tree), and the protein maximum likelihood tree, which places clade 3 with trypanosomatids (hereafter the protein tree). With each data set, we calculated the likelihoods for both trees considering first and second positions together, using a GTR+Γ+I model and empirical base frequencies, with all parameters optimized on the 1+2nuc tree. As expected, the 1+2nuc tree always had a higher likelihood than did the protein tree. We also scored the two trees under a more complex separate model where the parameters used at first- and second-codon positions were independent: we used the same GTR+Γ+I model and optimization procedures as before but considered only the first-codon positions and then repeated the process with second positions only. For each tree, the −lnL scores for the two data partitions were summed to obtain the likelihood for all the data.

In every case, the likelihoods for each tree are much higher under the separate model of sequence evolution (table 3). The likelihood ratios (ΔlnL) are highly significant (likelihood ratio tests; $P \ll 0.0001$, χ^2 df = 35–45, including the extra branch lengths). Most importantly, using a separate model affects the relative likelihoods of the two test trees; with the pK data set, the protein tree actually has a slightly higher likelihood than the 1+2nuc tree (table 3). With the other two data sets, the 1+2nuc tree remains more likely, but the likelihood ratio is reduced (despite the examination being slightly biased in favor of the 1+2nuc tree, under which the model parameters were optimized).

Inspection confirms large differences in the optimized substitution matrices for first and second positions under our separate model, with up to 10-fold differences in the relative rates estimated for the same substitution in the two data partitions (table 2). We performed a second analysis to separate the effect of the substitution matrices from the other parameters examined. Using the same three data sets we determined the likelihoods of the 1+2nuc and protein trees under a single GTR+Γ+I model. This time we calculated the likelihoods of first positions and second positions separately, then summed the −lnL scores. This allows different branch lengths for first and second positions. We then fixed the base frequency and among-site rate heterogeneity parameters, but optimized separate substitution matrices for first and second positions, and summed the −lnL scores for first and second positions for each tree. As with our former analysis, the single substitution ma-

Table 3
Likelihoods of the Two hsp90 Maximum Likelihood Tree Topologies Under Different Treatments of First and Second Codon Position Nucleotide Data

Analysis	Dataset	Treatment of 1 + 2 positions	1+2nuc tree (−lnL)	Prot tree (−lnL)	Δ lnL
All parameters reestimated.	pK	Combined	5047.895	5049.386	−1.491
		Separate	4933.709	4932.037	1.672*
	pKDHP	Combined	5974.239	5981.57	−7.331
		Separate	5849.892	5854.108	−4.216
	pKD	Combined	7755.55	7762.875	−7.325
		Separate	7593.418	7597.209	−3.791
Substitutions and branch lengths only . .	pK	Combined	5020.801	5022.435	−1.634
		Separate	4988.833	4987.708	1.125*
	pKDHP	Combined	5940.721	5947.599	−6.878
		Separate	5904.06	5908.794	−4.733
	pKD	Combined	7707.021	7714.542	−7.521
		Separate	7648.269	7653.597	−5.328

NOTE.—When separate parameter values are estimated for first and second positions, the superiority of the 1+2 nuc tree is reduced, and in the pK analysis, it becomes less likely than the protein tree (asterisks).

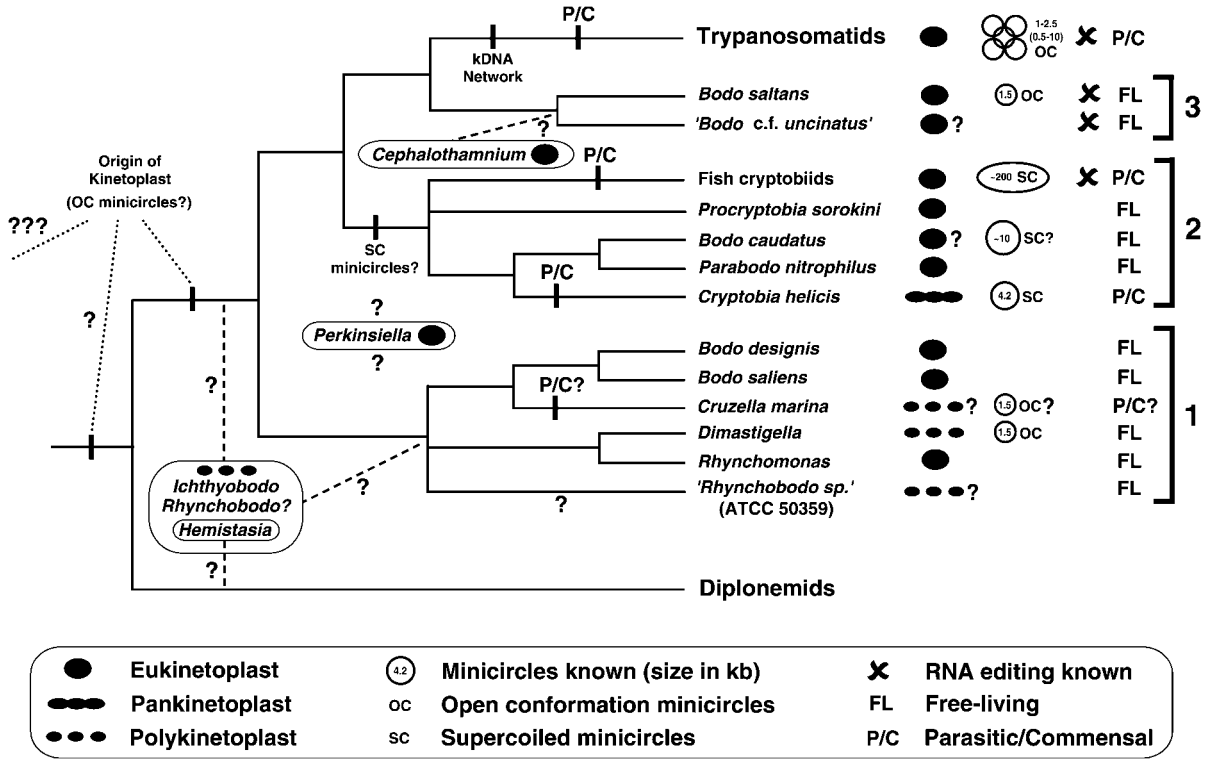

FIG. 4.—Model of kinetoplastid phylogeny based on the results of this study. Taxa for which there are molecular data are presented at terminal nodes. Other interesting taxa are listed in boxes within the tree, with their possible phylogenetic positions signified by dashed lines. The division of bodonids into three major clades (1–3) is shown at the far right. Important phenotypic features are signified by icons, as described in the legend at the bottom of the figure. Major inferred evolutionary events (origins of kinetoplast features, origins of parasitism-commensalism) are depicted on tree branches.

trix model always assigns a higher likelihood to the 1+2nuc tree (table 3). Allowing separate substitution matrices for first and second positions again significantly increases the likelihood of the data (likelihood ratio test, under a χ^2 distribution with df = 5, $P \ll 0.0001$). Again, the protein tree has a higher likelihood than that of the 1+2nuc tree with the pK data set, whereas the likelihood ratio shrinks with the other data sets. We conclude that large differences in sequence evolution at first- and second-codon positions, specifically, the relative rates of different substitutions, could be contributing to an incorrect phylogenetic estimation with the 1+2 nucleotide data. We therefore give more weight to our protein and 2-only nucleotide topology, which places clade 3 as the sister to trypanosomatids rather than as sister to clade 2.

Discussion

Our Model for the Relationships Among Kinetoplastids

Overview

Our study provides a new model for the evolutionary relationships within kinetoplastids (fig. 4). The known diversity of bodonids can be viewed as falling into three primary clades. Of these, clade 3 (*B. saltans* and *Bodo* cf. *uncinatus*) is most closely related to trypanosomatids. The basal split among known kinetoplastids is between clade 1 and all other taxa.

The Position of Trypanosomatids

Our hsp90 phylogenies provide the strongest phylogenetic evidence to date that *B. saltans* (with *Bodo* cf. *uncinatus*) may be a particularly close relative of trypanosomatids. This phylogenetic relationship was first suggested on the basis of similarity of the RNA editing of *CoxII* genes and a parsimony analysis of 296 bases of the edited *CoxII* gene transcript (Blom et al. 1998). In both these inferences, *C. helicis* and *T. borreli* were the only other bodonids considered (i.e., the sampling of bodonids was very limited). In fact, when *B. saltans* was included in SSUrRNA phylogenies (with many more bodonids represented, longer sequences, and, usually, with more sophisticated phylogenetic methods) it often did not form a clade with trypanosomatids: published SSUrRNA trees where *B. saltans* form a specific clade with trypanosomatids are preliminary results without robustness measures (Simpson et al. 2000) or have reduced taxon sampling and provide weak support (Lukeš et al. 2002). More detailed analyses from similar data sets actually place nodes with strong bootstrap support between trypanosomatids and *B. saltans-Bodo* cf. *uncinatus* (Simpson and Maslov 1999; Doležel et al. 2000).

The phylogenetic position of *B. saltans* makes this a key organism for understanding the evolution of the medically important trypanosomatids. Assuming that reversals to a free-living habitat from a parasitic ancestry

are rare, it seems likely that trypanosomatids acquired the parasitic condition independently of any of the parasitic bodonids that have been studied to date (fig. 4). Examinations of *B. saltans* could help unravel which of the unusual features of trypanosomatids might be associated with the adoption of parasitism and which appear to predate this evolutionary event.

The Root of the Kinetoplastid Tree

Our analysis of hsp90 data provides the most robust estimate to date for the position of the root of the kinetoplastid tree. A position between clade 1 and all other kinetoplastids is returned by almost all analyses, even though statistical support is weak with some taxon-method combinations, and some second-position nucleotide analyses (very marginally) favor an alternative placement. In contrast, previous studies of SSUrRNA data (the only published examinations that include taxa from all four of our clades) suggest many different placements for the root of kinetoplastids. Some published trees place the root within clade 2 (Maslov, Yasuhira, and Simpson 1999; Simpson and Maslov 1999; Lukeš et al. 2002), whereas others place the root inside clade 1, either with *Rhynchobodo* or with some or all of the *B. designis-B. saliens-C. marina* grouping at the base (Atkins, Teske, and Anderson 2000; Doležel et al. 2000). These rooting positions are never strongly supported in bootstrap analyses and can vary depending on the phylogenetic method used (Doležel et al. 2000). These previous studies mirror our SSUrRNA analyses, where out-group and method selection had a substantial effect on the placement of the root, with no one position ever being well supported.

It is suspected that the extreme length of the branch at the base of kinetoplastids in SSUrRNA analyses confounds the accurate placement of the root, even when relatively sophisticated models of sequence evolution are used. With the current taxon sampling, robustly placing the root of kinetoplastids with SSUrRNA might well be intractable. It is questionable whether anything is gained by including out-groups in SSUrRNA analyses of kinetoplastids. In contrast, with hsp90, the branch leading to kinetoplastids is much shorter in both proportionate and absolute terms, irrespective of the out-group taxa and phylogenetic methods used. The hsp90 data seem to be a more reliable foundation for inferences about the position of the root of kinetoplastids.

The Placement of Rhynchobodo

Within our model, the placement of one taxon, *Rhynchobodo*, remains somewhat uncertain. *Rhynchobodo* sp. ATCC 50359 appear as a member of clade 1 in SSUrRNA analyses but falls as the basal-most taxon in that clade with some taxon-method combinations (see also Doležel et al. 2000). If this basal placement is correct, our hsp90 results cannot rule out a rooting of the kinetoplastid tree between *Rhynchobodo* sp. and (the rest of) clade 1 because we have no hsp90 sequence from *Rhynchobodo* sp.

We also note that the generic assignment of *Rhynchobodo* sp. ATCC 50359 might be questionable. The taxon *Rhynchobodo* has only recently achieved some rationality (Bernard, Simpson, and Patterson 2000), and smaller members could still be confused with organisms assignable to *Bodo*. Discriminating morphological data are not available for ATCC 50359 (Doležel et al. 2000 allude to electron microscopical data, but this appears to be an error—A. P. Mylnikov, personal communication). Confirmations of both the identity and phylogenetic position of ATCC 50359 would be timely.

Have We Explored All of Kinetoplastid Diversity?

From the molecular perspective, the taxon sampling of bodonid kinetoplastids has improved drastically over recent years (compare Lukeš et al. 1998 to Doležel et al. 2000). However, there remain some known groups for which there is no sequence information, including the ectocommensal *Cephalothamnium*, the fish ectoparasite *Ichthyobodo*, the free-living *Hemistasia*, and a kinetoplastid isolated from the blowhole of a whale, *Jarrellia* (Poynton, Whitaker, and Heinrich 2001), although we suspect that the latter organism may belong to one of the known cryptobiid clades. This list might also include genuine *Rhynchobodo* (if *Rhynchobodo* sp. ATCC 50359 is misidentified—see above) and *Perkinsiella*, a eukinetoplast-bearing symbiont of certain amoebae (Hollande 1980; Dyková, Figueras, and Peric 2000). *Cephalothamnium* is morphologically rather similar to *B. saltans* (Hitchen 1974) and may be important for understanding the origin of trypanosomatids. *Ichthyobodo* and *Hemistasia* (and *Rhynchobodo*) are polykinetoplast-bearing organisms (Joyon and Lom 1969; Vørs 1992; Elbrächter, Schnepf, and Balzer 1996; Bernard, Simpson, and Patterson 2000). The polykinetoplastidic *Dimastigella*, *Cruzella*, and (probably) *Rhynchobodo* sp., all belong to the early diverging clade 1. It is possible that other polykinetoplastidic organisms occupy basal positions in the kinetoplastid tree. Of particular interest is *Hemistasia* because the scanty ultrastructural data available (Elbrächter, Schnepf, and Balzer 1996) hint that this organism may be most closely related to diplonemids (Simpson 1997; A. G. B. Simpson, unpublished data). This would imply either that the putative kinetoplast of *Hemistasia* is not homologous to that of other kinetoplastids or that the origin of the kinetoplast predates the split between diplonemids and typical kinetoplastids (note also the recent report of kDNA-like material in the mitochondrion of an euglenid; Leander, Triemer, and Farmer 2001). Resolving the evolutionary position of *Hemistasia* may be particularly important for uncovering the evolutionary history of kinetoplastids and the kinetoplast.

The Systematics of Kinetoplastids

Our study builds on recent explorations indicating that many widely used taxa within kinetoplastids are not monophyletic (Doležel et al. 2000; Maslov, Podlipaev, and Lukeš 2001). We provide the most robust evidence to date that Bodonina (=bodonids, as used here) is a

paraphyletic assemblage. In fact, as far as we are aware, all taxa used in the last 30 years to group multiple bodonid genera together are robustly nonmonophyletic in both our SSUrRNA and hsp90 phylogenies. This includes Bodonidae (often identical in composition to Bodonina), Bodoninae, Cryptobiidae, and Cryptobiinae (see Vickerman 1976, 1990; Zhukov 1991).

A major revision of kinetoplastid systematics is required. For the present, we recommend abandoning Bodonina, Bodonidae, and any other formal taxa corresponding to bodonids. These should be ''no regrets'' actions, at least for phylogenetic systematists and their sympathizers. Providing further research confirms their monophyly, we can imagine major new taxa within kinetoplastids corresponding to clade 1 and clade 2, and reflecting the close relationship between trypanosomatids and clade 3, although we are unable to suggest unambiguous structural apomorphies for any of these groups at present. We note that the latter taxon could not be included in a system that also recognizes Bodonina.

On the Lack of Introns

This is the first study wherein any nuclear-encoded gene has been sequenced from a wide diversity of kinetoplastids. Outside trypanosomatids, publicly available data appear restricted to the clade 2 *T. borreli*, although we have begun to determine cytoplasmic heat shock protein 70 (hsp70) gene sequences from some other bodonids (A. G. B. Simpson, unpublished data). None of these hsp90 genes or other nuclear sequences from bodonids contain introns (whereas hsp90 genes from other Excavata, even the intron-poor *N. gruberi*, usually do have introns—this study; A. G. B. Simpson, unpublished data, and nuclear genes from euglenids can be intron-rich—e.g., Canaday et al. 2001). Trypanosomatids have remarkably compact nuclear genomes, including fairly short intergenic spacers with polycistronic transcription (Myler et al. 1999), overlapping genes (Liniger et al. 2001), and an extremely low abundance of spliceosomal introns, such that despite major genomic efforts, these elements have been discovered only recently (Mair et al. 2000). It is tempting to interpret this streamlining as an evolutionary response to the adoption of a parasitic habit. However, our survey hints that introns might be rare in both parasitic and free-living kinetoplastids and that this general condition could predate the (probably comparatively late) adoption of parasitism by the trypanosomatid lineage. Certainly, this is an important possibility deserving further exploration.

Kinetoplast Evolution

This study influences our understanding of the evolution of kinetoplast DNA (kDNA). The suggested position for the root of kinetoplastids embeds trypanosomatids deep within bodonids in general, and also within the subset of bodonids for which there are data on kDNA or RNA editing (or both). This confirms that examinations of a diversity of bodonids will assist in charting the evolution of the well-studied trypanosomatid systems. However, our study also demonstrates that the set of organisms for which kDNA or RNA editing (or both) data are available do not well represent kinetoplastids as a whole, being skewed toward fairly close relatives of trypanosomatids (clades 2 and 3—see fig. 4). There is a single preliminary published account of minicircle organization from clade 1 organisms (Štolba, Jirků, and Lukeš 2001), and there is no RNA editing information at all. A greater emphasis on these early diverging kinetoplastids (relative to trypanosomatids) will be required for a comprehensive understanding of kDNA evolution.

Considering minicircle organization, data are published for the clade 3 *B. saltans*, several clade 2 organisms, and the clade 1 *Dimastigella*, whereas the clade 1 *Cruzella* apparently has a similar minicircle organization to *Dimastigella* (A. Zíková, M. Vancova, and J. Lukeš, personal communication; fig. 4). These data do allow some evolutionary inference. Because all these bodonids have mostly noncatenated minicircles, our analysis confirms that the true network is a derived condition for trypanosomatids (Simpson et al. 2000; Lukeš et al. 2002). The proximity of *B. saltans* and trypanosomatids may tempt the inference that their open-conformation minicircles are a shared-derived condition (Lukeš et al. 2002). In fact, the kinetoplastids that are believed to have supercoiled minicircles or megacircles (*B. caudatus*, *C. helicis*, *T. borreli*) all belong to clade 2. Clade 1 organisms for which there are data (*Dimastigella*, *Cruzella*) have open-conformation minicircles, similar in size to those in *B. saltans* (Štolba, Jirků, and Lukeš 2001; A. Zíková, M. Vancova and J. Lukeš, personal communication). Even assuming that open-conformation minicircles are derived from some form of supercoiled circular DNA, our phylogenetic model suggests two scenarios that are equally consistent under naïve parsimony—either (1) open-conformation minicircles are the ancestral condition for studied kinetoplastids, with a reversion to supercoiling in clade 2 or (2) the open-conformation condition was independently acquired in clade 1 and in an ancestor of clade 3 + trypanosomatids. Given the rarity of open-conformation DNA circles in nature, the former scenario seems more possible. We therefore suggest that open-conformation minicircles may have arisen early in kinetoplastid evolution, substantially predating the origin of the true interlocking network. Certainly, there is no a priori reason why the possession of one derived condition by trypanosomatids (the network) should mean that other aspects of their kDNA structure (open-conformation minicircles) are also derived states.

Supplementary Information

The GenBank accession numbers for the *hsp90* gene sequences reported in this study are AY122622–AY122634.

After the review of this paper, Callahan, Litaker, and Noga (2002) reported several new SSUrRNA sequences from bodonids, including the first from *Ichthyobodo*. The *Ichthyobodo* sequence appears to be very

divergent and is strongly attracted to the out-group in the published phylogenetic analysis. Its position relative to other kinetoplastids is highly uncertain when out-groups are omitted (H. A. Callahan, personal communication; A. G. B. Simpson, unpublished data). Based on their SSUrRNA trees, Callahan, Litaker, and Noga (2002) denote a number of nonnested sublineages within kinetoplastids (excluding *Ichthyobodo*). Their sublineage 4 corresponds to our clade 3. Otherwise, their sublineages 2–4 subdivide our clade 2 and their 5–8 subdivides our clade 1. They also point out the nonmonophyly of Bodonidae and of Cryptobiidae. Interestingly, their SSUrRNA analysis also indicates a sistergroup relationship between clade 3 and trypanosomatids, although, again (probably) with weak bootstrap support.

Acknowledgments

We would like to thank M. Atkins for providing genomic DNA from *B. saliens* and *R. nasuta*, and J. D. Silberman and J. B. Dacks for critical comments. This research was supported by Grant 227085-00 awarded to A.J.R. by the Natural Sciences and Engineering Research Council of Canada. A.G.B.S. is supported by a Canadian Institutes of Health Research Postdoctoral Fellowship. A.J.R. thanks the Canadian Institute for Advanced Research Program in Evolutionary Biology for fellowship support.

LITERATURE CITED

ARNDT, H., D. DIETRICH, B. AUER, E. CLEVEN, T. GRAFENHAM, M. WEITERE, and A. P. MYLNIKOV. 2000. Functional diversity of heterotrophic flagellates in aquatic ecosystems. Pp. 240–268 *in* B. S. C. LEADBEATER and J. C. GREEN, eds. The flagellates. Taylor and Francis, London.

ATKINS, M. S., A. P. TESKE, and O. R. ANDERSON. 2000. A survey of flagellate diversity at four deep-sea hydrothermal vents in the eastern Pacific Ocean using structural and molecular approaches. J. Eukaryot. Microbiol. **47**:400–411.

BALDAUF, S. L., A. J. ROGER, I. WENK-SIEFERT, and W. F. DOOLITTLE. 2000. A kingdom-level phylogeny of eukaryotes based on combined protein data. Science **290**:972–977.

BERNARD, C., A. G. B. SIMPSON, and D. J. PATTERSON. 2000. Some free-living flagellates (Protista) from anoxic habitats. Ophelia **52**:113–142.

BLOM, D., A. DE HAAN, M. VAN DEN BERG, P. SLOOF, M. JIRKŮ, J. LUKEŠ, and R. BENNE. 1998. RNA editing in the free-living bodonid *Bodo saltans*. Nucleic Acids Res. **26**: 1205–1213.

BLOM, D., A. DE HAAN, J. VAN DEN BURG, M. VAN DEN BERG, P. SLOOF, M. JIRKŮ, J. LUKEŠ, and R. BENNE. 2000. Mitochondrial minicircles in the free-living bodonid *Bodo saltans* contain two gRNA gene cassettes and are not found in large networks. RNA **6**:121–135.

CALLAHAN, H. A., R. W. LITAKER, and E. J. NOGA. 2002. Molecular taxonomy of the suborder Bodonina (Order Kinetoplastida), including the important fish parasite, *Ichthyobodo necator*. J. Eukaryot. Microbiol. **49**:119–128.

CANADAY, J., L. H. TESSIER, P. IMBAULT, and F. PAULUS. 2001. Analysis of *Euglena gracilis* alpha-, beta- and gamma-tubulin genes: introns and pre-mRNA maturation. Mol. Genet. Genomics **265**:153–160.

DOLEŽEL, D., M. JIRKŮ, D. A. MASLOV, and J. LUKEŠ. 2000. Phylogeny of the bodonid flagellates (Kinetoplastida) based on small-subunit rRNA gene sequences. Int. J. Syst. Evol. Microbiol. **50**:1943–1951.

DONELSON, J. E., M. J. GARDNER, and N. M. EL-SAYED. 1999. More surprises from Kinetoplastida. Proc. Natl. Acad. Sci. USA **96**:2579–2581.

DYKOVÁ, I., A. FIGUERAS, and Z. PERIC. 2000. *Neoparamoeba* Page, 1987: light and electron microscopic observations on six strains of different origin. Dis. Aquat. Org. **43**:217–223.

ELBRÄCHTER, M., E. SCHNEPF, and I. BALZER. 1996. *Hemistasia phaeocysticola* (Scherffel) comb. nov., redescription of a free-living, marine phagotrophic kinetoplastid flagellate. Arch. Protistenkd. **147**:125–136.

ESTÉVEZ, A. M., and L. SIMPSON. 1999. Uridine insertion/deletion RNA editing in trypanosome mitochondria: a review. Gene **240**:247–260.

FELSENSTEIN, J. 1995. PHYLIP (phylogeny inference package). Version 3.57c. Distributed by the author, University of Washington, Seattle.

———. 2000. PHYLIP (phylogeny inference package). Version 3.6a. Distributed by the author, University of Washington, Seattle.

HAJDUK, S. L., A. M. SIQUEIRA, and K. VICKERMAN. 1986. Kinetoplast DNA of *Bodo caudatus*: a noncatenated structure. Mol. Cell. Biol. **6**:4372–4378.

HITCHEN, E. T. 1974. The fine structure of the colonial kinetoplastid flagellate *Cephalothamnium cyclopum*. J. Protozool. **21**:221–231.

HOLLANDE, A. 1980. Identification du parasome (Nebenkern) de *Janickina pigmentifera* aun symbionte (*Perkinsiella amoebae* nov gen, nov sp.) apparenté aux flagellés kinetoplastidiés. Protistologica **16**:613–625.

JOYON, L., and J. LOM. 1969. Etude cytologique, systématique et pathologique d'*Ichtyobodo necator* (Henneguy, 1883) Pinto, 1928 (zooflagelle). J. Protozool. **16**:705–719.

KIVIC, P. A., and P. L. WALNE. 1984. An evaluation of a possible phylogenetic relationship between the Euglenophyta and Kinetoplastida. Origins Life **13**:269–288.

KLINGBEIL, M. M., M. E. DREW, Y. LIU, J. C. MORRIS, S. A. MOTYKA, T. T. SAXOWSKY, Z. WANG, and P. T. ENGLUND. 2001. Unlocking the secrets of trypanosome kinetoplast DNA network replication. Protist **152**:255–262.

LANDWEBER, L. F., and W. GILBERT. 1994. Phylogenetic analysis of RNA editing: a primitive genetic phenomenon. Proc. Natl. Acad. Sci. USA **91**:918–921.

LEANDER, B. S., R. E. TRIEMER, and M. A. FARMER. 2001. Character evolution in heterotrophic euglenids. Eur. J. Protistol. **37**:337–356.

LINIGER, M., K. BODENMÜLLER, E. PAYS, S. GALLATI, and I. RODITI. 2001. Overlapping sense and antisense transcription units in *Trypanosoma brucei*. Mol. Microbiol. **40**:869–878.

LUKEŠ, J., G. J. ARTS, M. VAN DEN BERG, A. DE HAAN, F. R. OPPERDOES, P. SLOOF, and R. BENNE. 1994. Novel pattern of editing regions in mitochondrial transcripts of the cryptobiid *Trypanoplasma borreli*. EMBO J. **13**:5086–5098.

LUKEŠ, J., J. L. GUILBRIDE, J. VOTÝPKA, A. ZÍKOVÁ, R. BENNE, and P. T. ENGLUND. 2002. The kinetoplast DNA network: evolution of an improbable structure. Eukaryot. Cell **1**:495–502.

LUKEŠ, J., M. JIRKŮ, N. AVLIYAKULOV, and O. BENADA. 1998. Pankinetoplast DNA structure in a primitive bodonid flagellate, *Cryptobia helicis*. EMBO J. **17**:838–846.

MAIR, G., H. SHI, H. LI et al. 2000. A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. RNA **6**: 163–169.

MASLOV, D. A., H. A. AVILA, J. A. LAKE, and L. SIMPSON. 1994. Evolution of RNA editing in kinetoplastid protozoa. Nature **368**:345–348.

MASLOV, D. A., S. A. PODLIPAEV, and J. LUKEŠ. 2001. Phylogeny of the Kinetoplastida: taxonomic problems and insights into the evolution of parasitism. Mem. Inst. Oswaldo Cruz **96**:397–402.

MASLOV, D. A., and L. SIMPSON. 1994. RNA editing and mitochondrial genomic organization in the cryptobiid kinetoplastid protozoan *Trypanoplasma borreli*. Mol. Cell. Biol. **14**:8174–8182.

MASLOV, D. A., S. YASUHIRA, and L. SIMPSON. 1999. Phylogenetic affinities of *Diplonema* within the Euglenozoa as inferred from the SSU rRNA gene and partial CO1 protein sequences. Protist **150**:33–42.

MOREIRA, D., P. LOPEZ-GARCIA, and F. RODRIGUEZ-VALERA. 2001. New insights on the phylogenetic position of diplonemids: GC content bias, differences of evolutionary rate, and a new environmental sequence. Int. J. Syst. Evol. Microbiol. **51**:2211–2219.

MYLER, P., L. AUDLEMAN, T. DEVOS et al. 1999. *Leishmania major* Friedlin chromosome 1 has an unusual distribution of protein-coding genes. Proc. Natl. Acad. Sci. USA **96**: 2902–2906.

PEARL, L. H., and C. PRODROMOU. 2000. Structure and in vivo function of hsp90. Curr. Opin. Struct. Biol. **10**:46–51.

POYNTON, S., B. WHITAKER, and A. HEINRICH. 2001. A novel trypanoplasm-like flagellate *Jarrellia atramenti* n. g., n. sp. (Kinetoplastida: Bodonidae) and ciliates from the blowhole of a stranded pygmy sperm whale *Kogia breviceps* (Physeteridae): morphology, life cycle and potential pathogenicity. Dis. Aquat. Org. **44**:191–201.

SIMPSON, A. G. B. 1997. The identity and composition of the Euglenozoa. Arch. Protistenkd. **148**:318–328.

SIMPSON, L. 1997. The genomic organisation of guide RNA genes in kinetoplastid protozoa: several conundrums and their solutions. Mol. Biochem. Parasitol. **86**:133–141.

SIMPSON, L., and D. A. MASLOV. 1999. Evolution of the U-insertion/deletion RNA editing in mitochondria of kinetoplastid protozoa. Ann. N. Y. Acad. Sci. **870**:190–205.

SIMPSON, L., O. H. THIEMANN, N. J. SAVILL, J. D. ALFONZO, and D. A. MASLOV. 2000. Evolution of RNA editing in trypanosome mitochondria. Proc. Natl. Acad. Sci. USA **97**: 6986–6993.

ŠTOLBA, P., M. JIRKŮ, and J. LUKEŠ. 2001. Polykinetoplast DNA structure in *Dimastigella trypaniformis* and *Dimastigella mimosa* (Kinetoplastida). Mol. Biochem. Parasitol. **113**:323–326.

STRIMMER, K., and A. VON HAESELER. 1996. Quartet puzzling: a quartet maximum-likelihood method for reconstructing tree topologies. Mol. Biol. Evol. **13**:964–969.

SWOFFORD, D. L. 2000. PAUP*: phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, Mass.

VICKERMAN, K. 1976. The diversity of the kinetoplastid flagellates. Pp. 1–34 *in* W. LUMSDEN and D. EVANS, eds. Biology of the Kinetoplastida. Academic Press, London.

———. 1977. DNA throughout the single mitochondrion of a Kinetoplastid flagellate: observations on the ultrastructure of *Cryptobia vaginalis* (Hesse, 1910). J. Protozool. **24**:221–233.

———. 1990. Phylum Zoomastigina, Class Kinetoplastida. Pp. 215–238 *in* L. MARGULIS, J. O. CORLISS, M. MELKONIAN, and D. J. CHAPMAN, eds. Handbook of Protoctista. Jones and Bartlett, Boston.

VØRS, N. 1992. Heterotrophic amoebae, flagellates and heliozoa from the Tvärminne area, Gulf of Finland, in 1988–1990. Ophelia **36**:1–109.

WILGENBUSCH, J., and K. DE QUEIROZ. 2000. Phylogenetic relationships among the phrynosomatid sand lizards inferred from mitochondrial DNA sequences generated by heterogeneous evolutionary processes. Syst. Biol. **49**:592–612.

WOO, P. T. K. 2001. Cryptobiosis and its control in North American fishes. Int. J. Parasitol. **31**:566–574.

WRIGHT, A.-D. G., S. LI, S. FENG, D. S. MARTIN, and D. H. LYNN. 1999. Phylogenetic position of the kinetoplastids, *Cryptobia bullocki*, *Cryptobia catostomi*, and *C. salmositica* and monophyly of the genus *Trypanosoma* inferred from small subunit ribosomal RNA sequences. Mol. Biochem. Parasitol. **99**:69–76.

YASUHIRA, S., and L. SIMPSON. 1996. Guide RNAs and guide RNA genes in the cryptobiid kinetoplastid protozoan, *Trypanoplasma borreli*. RNA **2**:1153–1160.

ZHUKOV, B. 1991. The diversity of bodonids. Pp. 177–184 *in* D. PATTERSON and J. LARSEN, eds. The biology of free-living heterotrophic flagellates. Clarendon Press, Oxford.

FRANZ LANG, reviewing editor

Accepted June 11, 2002