Phylogenomic Analysis of Kinetoplastids Supports That Trypanosomatids Arose from within Bodonids

Philippe Deschamps,1 Enrique Lara,1,2 William Marande,1,3 Purificación López-García,1 Flemming Ekelund,4 and David Moreira*.1
1Unité d’Ecologie, Systématique et Evolution, UMR CNRS 8079, Univ. Paris-Sud, 91405 Orsay, France
2Laboratory of Soil Biology, Institute of Biology, University of Neuchâtel, Neuchâtel, Switzerland
3FRE 3206 CNRS-Muséum National d’Histoire Naturelle, Paris, France
4Department of Biology, University of Copenhagen, Copenhagen, Denmark
*Corresponding author: E-mail: david.moreira@u-psud.fr.
Associate editor: Charles Delwiche

Abstract

Kinetoplastids are a large group of free-living and parasitic eukaryotic flagellates, including the medically important trypanosomatids (e.g., Trypanosoma and Leishmania) and the widespread free-living and parasitic bodonids. Small subunit rRNA- and conserved protein-based phylogenies support the division of kinetoplastids into five orders (Prokinetoplastida, Neobodonida, Parabodonida, Eubodonida, and Trypanosomatida), but they produce incongruent results regarding their relative branching order, in particular for the position of the Trypanosomatida. In general, small subunit rRNA tends to support their early emergence, whereas protein phylogenies most often support a more recent origin from within bodonids. In order to resolve this question through a phylogenomic approach, we carried out massive parallel sequencing of cDNA from representatives of three bodonid orders (Bodo saltans -Eubodonida-, Procryptobia sorokini - Parabodonida-, and Rhynchomonas nasuta -Neobodonida-). We identified 64 well-conserved proteins shared by these species, four trypanosomatids, and two closely related outgroup species (Euglena gracilis and Diplonema papillatum). Phylogenetic analysis of a concatenated data set yielded a strongly supported tree showing the late emergence of trypanosomatids as a sister group of the Eubodonida. In addition, we identified homologues of proteins involved in trypanosomatid mitochondrial mRNA editing in the three bodonid species, suggesting that editing may be widespread in kinetoplastids. Comparison of expressed sequences from mitochondrial genes showed variability at U positions, in agreement with the existence of editing activity in the three bodonid orders most closely related to trypanosomatids (Neobodonida, Parabodonida, and Eubodonida). Mitochondrial mRNA editing appears to be an ancient phenomenon in kinetoplastids.

Key words: kinetoplastids, phylogenomics, eukaryotic phylogeny, editing.
Deschamps et al. · doi:10.1093/molbev/msq289

Fig. 1 ML phylogenetic tree of kinetoplastids reconstructed upon a concatenation of 64 conserved proteins and rooted on Euglena gracilis and Diplonema papillatum as outgroups. Numbers at nodes are BI posterior probabilities and ML bootstrap values.

was very variable, depending on the taxon sampling and analysis methods, and in general remained moderate. Thus, efforts to improve the sampling of protein data sets for these organisms have been qualified as highly desirable (Simpson et al. 2004).

In addition to their medical importance, trypanosomatids have extreme versions of several of the very unusual characters found in kinetoplastids, such as mitochondrial RNA editing, generalized mRNA trans-splicing, or a complex mitochondrial genome structure (Simpson et al. 2006). Therefore, the resolution of their phylogenetic position is important to better understand the evolution of these exotic traits (Simpson et al. 2006). Discrepancies between the SSU rDNA and protein-based phylogenetic analyses are most likely due to the limited amount of phylogenetic signal contained in these markers and probably also to the fast evolutionary rate affecting these organisms. This is particularly noticeable for the SSU rDNA, which yields phylogenies with a very long basal branch that reflects a strong rate acceleration early in the evolution of kinetoplastids (Dolezel et al. 2000; Simpson et al. 2002, 2004; Moreira et al. 2004; von der Heyden et al. 2004). To address this question, we tried to apply a multimarker approach, which is becoming the preferred choice to resolve difficult nodes using as queries a set of 143 protein sequences commonly employed in eukaryotic phylogenomic analyses (Rodríguez-Ezepeleta et al. 2005; Burki et al. 2007; Rodríguez-Ezepeleta, Brinkmann, Burger, et al. 2007; Hampl et al. 2009). We also looked for homologues of the 178 proteins (most of them hypothetical) identified in a pilot study of the genome of B. saltans (Jackson et al. 2008). Among all the homologues retrieved from the cDNA sequences, we kept only those proteins found in the three bodonid samples and also in trypanosomatids and in at least one closely related outgroup species (Euglena gracilis and/or Diplonema papillatum, Marande and Burger 2007; Rodríguez-Ezepeleta, Brinkmann, Burger, et al. 2007). In addition, we excluded those proteins that after multiple sequence alignment did not show any overlapping region for the three bodonids due to partial sequencing of the corresponding cDNAs. This yielded a final list of 64 proteins (supplementary table 1, Supplementary Material online).

We first carried out individual phylogenetic analyses of these proteins with a large eukaryotic taxon sampling to check for potential hidden paralogy or horizontal gene transfer cases. All of them retrieved the monophyly of kinetoplastids and did not show any sign of these problems. Hence, we analyzed them simultaneously in a concatenated data set with only the kinetoplastid + close outgroup species (E. gracilis and D. papillatum) in order to maximize the number of conserved positions and to avoid potential artifacts due to the use of more distant outgroup taxa (Rodríguez-Ezepeleta, Brinkmann, Roure, et al. 2007). The resulting data set contained 11,425 amino acids and a maximum of 31% missing data for the kinetoplastid sequences (supplementary table 2, Supplementary Material online). Maximum likelihood (ML), with the LG + Γ model (Le and Gascuel 2008), and Bayesian inference (BI), with the mixture model CAT (Lartillot and Philippe 2004), phylogenetic analyses of this data set yielded the same, strongly supported tree (fig. 1). It showed the late emergence of trypanosomatids as sisters of the free-living bodonid B. saltans, as already found in several single-protein analyses (Simpson et al. 2002, 2004) but, in contrast to the moderate support found in most of them, this node received maximal statistical support (ML bootstrap value—BV—of 100% and BI posterior probability—PP—of 1). We obtained the same results from the analysis of a data set with all missing data removed (supplementary fig. 1, Supplementary Material online). The SSU rDNA phylogeny for the same set of species showed a poorly supported trypanosomatid-early topology (supplementary fig. 2, Supplementary Material online), as previously observed in taxon-rich SSU rDNA phylogenies (Moreira, López-García, and Vickerman 2004; von der Heyden et al. 2004). In order to evaluate the robustness of these different trees, we carried out approximately unbiased (AU) tests (Shimodaira 2002) using the concatenated protein and SSU rDNA data sets. In addition to the kinetoplastid-late and early topologies, we also evaluated all possible topologies for the kinetoplastid taxa, except those implying the nonmonophyly of the trypanosomatids (i.e., 12 topologies, see table 1). The
concatenated protein data set significantly rejected the SSU rDNA topology \((P = 0)\) and all other topologies, except one where \(P.\) sorokini and \(R.\) nasuta formed a clade at the base of the kinetoplastids \((P = 0.498)\). In fact, this relationship was retrieved in 14 of the trees reconstructed using the individual markers, although always with weak support \((BV < 70\%)\). Compared with the concatenated protein data set, the SSU rDNA data set was much less discriminant, as it only rejected 6 of the 12 topologies tested (table 1), which likely reflected the relatively weak phylogenetic signal provided by that single marker.

We evaluated the size of the sequence data set required to resolve the position of the trypanosomatids within the kinetoplastids by constructing cumulative concatenations of the 64 proteins, adding them progressively in random order and repeating this operation 50 times to calculate average bootstrap values (fig. 2). Although the monophyly of trypanosomatids rapidly received a BV of 100% from all concatenations \(>8\) proteins, their late emergence as sisters of \(B. \) saltans required the concatenation of at least 40 proteins to reach a BV \(> 95\%\). The deepest node in the kinetoplastid phylogeny, concerning the basal position of \(R.\) nasuta, was supported by BV \(> 95\%\) by concatenations >50 proteins, which suggested that this position was correct despite the fact that the alternative of the sister grouping of this species with \(P.\) sorokini was not rejected by an AU test upon the complete concatenated data set (see above). These results confirmed the need of a relatively large sequence data set to reconstruct the deep nodes of the kinetoplastid phylogeny.

As mentioned above, extensive mitochondrial mRNA editing is an astonishing process found in trypanosomatids. It is carried out through a complex series of insertions and deletions of uridylic residues (Us) with the help of guide RNAs to produce the mature and functional mRNAs (Blum et al. 1990; Alfonzo et al. 1997). Several proteins are involved, constituting complexes such as the 20S editosome (Worthy et al. 2003; Osato et al. 2009) and MRB1 (Hashimi et al. 2008, 2009; Acestor et al. 2009) and KPAP1 (Etheridge et al. 2008) complexes. Among bodonids, editing has only been proved in \(T.\) borreli (Lukes\' et al. 1994; Maslov and Simpson 1994) and \(B.\) saltans mitochondria (Blom et al. 1998). To look for possible evidence of editing in other free-living bodonids, we searched for homologues of the 20S editosome and MRB1 and KPAP1 complexes characterized in \(T.\) cruzi and \(L.\) major.

Table 1. Results of the AU Tests Using the Complete Protein (64 proteins) and SSU rDNA Data sets.

<table>
<thead>
<tr>
<th>Tree Topology</th>
<th>Proteins</th>
<th>SSU rDNA</th>
</tr>
</thead>
<tbody>
<tr>
<td>((Egra, (Dpap, (Psor, (Bsal, (Linf, Lmaj), (Tbru, Tcru)))))</td>
<td>0.968</td>
<td>0.334</td>
</tr>
<tr>
<td>((Egra, (Dpap, ((Bsal, Psor), ((Linf, Lmaj), (Tbru, Tcru)))))</td>
<td>0</td>
<td>0.871</td>
</tr>
<tr>
<td>((Egra, (Dpap, (Psor, (Bsal, Psor), ((Linf, Lmaj), (Tbru, Tcru)))))</td>
<td>0</td>
<td>0.313</td>
</tr>
<tr>
<td>((Egra, (Dpap, (Bsal, Psor, (Linf, Lmaj), (Tbru, Tcru)))))</td>
<td>0</td>
<td>0.120</td>
</tr>
<tr>
<td>((Egra, (Dpap, (Psor, (Bsal, Lmaj), (Tbru, Tcru)))))</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>((Egra, (Dpap, ((Linf, Lmaj), (Tbru, Tcru)), (Rnas, Psor))))</td>
<td>0.498</td>
<td>0.273</td>
</tr>
<tr>
<td>((Egra, (Dpap, (Psor, (Bsal, Psor), ((Linf, Lmaj), (Tbru, Tcru)))))</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>((Egra, (Dpap, (Bsal, (Linf, Lmaj), (Tbru, Tcru)), (Rnas, Psor))))</td>
<td>0.013</td>
<td>0</td>
</tr>
<tr>
<td>((Egra, (Dpap, (Psor, (Bsal, (Linf, Lmaj), (Tbru, Tcru)))))</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

*Note.*—Significant \(P\) values are in bold. \(Egra,\) *Euglena gracilis*; \(Dpap,\) Diplonema papillatum; \(Bsal,\) Bodo saltans; \(Psor,\) Procryptobia sorokini; \(Rnas,\) Rhynchomonas nasuta; \(Linf,\) *Leishmania infantum*; \(Lmaj,\) *Leishmania major*; \(Tbru,\) *Trypanosoma brucei*; \(Tcru,\) *Trypanosoma cruzi*.
species. Although we found very few hits against editing-related proteins in our cDNA sequences, likely due to the limited coverage of the cDNA sequencing, they were distributed in the three bodonid species (table 2). We found 7 occurrences in *P. sorokini*, 8 in *B. saltans*, and 5 in *R. nasuta*. In addition, homologues of the helicase Hel61 were found in the three bodonids but they were divergent and their orthology was unclear. Nevertheless, although a specific screening of editing-related genes will be necessary to verify the presence of functional editing complexes in the different bodonid groups, our results suggest that they may be present in species of the three bodonid orders of Metakinetoplastina. In addition to the presence of these genes, we looked for evidence of editing by examining the sequences of cDNAs of typical mitochondrial genes. We found homologues of *cox1*, *cox2*, and *nadh5* in our bodonid species showing a high proportion of U residues, which is characteristic of highly edited transcripts in kinetoplastids (Blum et al. 1990; Alfonzo et al. 1997). Moreover, we detected sequence variation at some positions corresponding to U residues, which has been used as evidence for editing in *T. borreli* (Lukáš et al. 1994; Maslov and Simpson 1994) and *B. saltans* (Blom et al. 1998). We observed a similar situation also in *P. sorokini* and *R. nasuta*, especially for the *cox2* gene, which paradoxically shows very limited editing in *B. saltans* (Blom et al. 1998) and no editing in *T. borreli* (Lukáš et al. 1994). Several *P. sorokini* and *R. nasuta* *cox2* transcripts showed several positions with a variable amount of missing U residues (fig. 3), probably corresponding to different stages in the editing process or to inefficient editing. This is particularly interesting for *R. nasuta* because it would be the first detection of editing in a member of the Neobodonida. These data and the presence of editing-related protein-coding genes support that editing, already demonstrated in *T. borreli* (Parabodonida) and *B. saltans* (Eubodonida) (Lukáš et al. 1994; Maslov and Simpson 1994), is widespread among bodonid species.

**Table 2.** Proteins Involved in Editing Found in the cDNA Sequences of Bodonid Species.

<table>
<thead>
<tr>
<th>Protein Complex</th>
<th>Bodo saltans</th>
<th>Procrpytobia sorokini</th>
<th>Rhynchosoma nasuta</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tb11.01.0610 (Hel61)</td>
<td>20S editosome</td>
<td>?</td>
<td>?</td>
</tr>
<tr>
<td>Tb11.03.0180 (MP44)</td>
<td>20S editosome</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Tb11.02.0490 (MP46)</td>
<td>20S editosome</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tb10.70.3850 (MP99)</td>
<td>20S editosome</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tb09.160.2970 (REL1)</td>
<td>20S editosome</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tb927.7.1550 (RET2)</td>
<td>20S editosome</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Tb10.406.0050</td>
<td>MRB1</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Tb927.2.3800 (GAP1)</td>
<td>MRB1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Tb927.5.3010</td>
<td>MRB1</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tb927.6.2140</td>
<td>MRB1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Tb927.7.2570 (GAP2)</td>
<td>MRB1</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tb10.389.0070 (EF-Tu)</td>
<td>KAP1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Tb11.01.7510</td>
<td>KAP1</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tb11.02.5820 (KPAP1)</td>
<td>KAP1</td>
<td>+</td>
<td>-</td>
</tr>
<tr>
<td>Tb11.47.0024</td>
<td>KAP1</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tb927.1.3010</td>
<td>KAP1</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tb927.2.3180 (TBPPR1)</td>
<td>KAP1</td>
<td>-</td>
<td>+</td>
</tr>
<tr>
<td>Tb927.4.4150</td>
<td>KAP1</td>
<td>+</td>
<td>-</td>
</tr>
</tbody>
</table>

*NOTE.—Reference sequences were from *T. cruzi*, (locus tag numbers are provided). +, a homologue was detected; -, no homologue was detected; ?, a homologue was detected, but sequence similarity was low. Only proteins with at least one homologue in the bodonid species are shown (for the complete list of proteins searched, see supplementary table 3, Supplementary Material online).*
1994; Blom et al. 1998), may be present also in Neobodo-
Hida and, therefore, be general in the whole subclass Meta-
kinetoplastina. Editing is likely an ancient phenomenon in
kinetoplastids.

In conclusion, our multiprotein phylogenetic analyses
provided strong support for the late emergence of the try-
panosomatids and the paraphyletic branching of the three
bodoniid groups within the Metakinetoplastina, overcom-
ing the contradictory results obtained up to now from the
analysis of small sequence data sets (SSU rDNA and a few
proteins). As already suggested (Simpson et al. 2004, 2006),
the sister relationship of Trypanosomatida with Eu bodoni-
dida makes B. saltans a key species to study the origin of the
parasitic trypanosomatids, deserving high priority for com-
plete genome sequencing (Jackson et al. 2008).

Supplementary Material
Detailed Materials and Methods and supplementary tables
1–3 and figures 1–2 are available at Molecular Biology and
Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
This work was supported by a grant from the French Agence Nationale de la Recherche (ANR JC05-44674 to
D.M). We thank Hervé Philippe for sharing his alignments
proteins). As already suggested (Simpson et al. 2004, 2006),
the sister relationship of Trypanosomatida with Eu bodoni-
dida makes B. saltans a key species to study the origin of the
parasitic trypanosomatids, deserving high priority for com-
plete genome sequencing (Jackson et al. 2008).

Supplementary Material
Detailed Materials and Methods and supplementary tables
1–3 and figures 1–2 are available at Molecular Biology and
Evolution online (http://www.mbe.oxfordjournals.org/).

Acknowledgments
This work was supported by a grant from the French Agence Nationale de la Recherche (ANR JC05-44674 to
D.M). We thank Hervé Philippe for sharing his alignments
proteins). As already suggested (Simpson et al. 2004, 2006),
the sister relationship of Trypanosomatida with Eu bodoni-
dida makes B. saltans a key species to study the origin of the
parasitic trypanosomatids, deserving high priority for com-
plete genome sequencing (Jackson et al. 2008).

References
Acetor N, Panigrahi AK, Carnes J, Ziková A, Stuart KD. 2009. The
MRB1 complex functions in kinetoplastid RNA processing. RNA
new higher level classification of eukaryotes with emphasis on
Alfonzo JD, Thiemann O, Simpson L. 1997. The mechanism of U
insertion/deletion RNA editing in kinetoplastid mitochondria.
Blom D, de Haan A, van den Berg M, Sloof G, Jirku M, Lukes J,
Benne R. 1998. RNA editing in the free-living bodonid Bodo
Blum B, Bakalara N, Simpson L. 1990. A model for RNA editing in
kinetoplastid mitochondria: “guide” RNA molecules transcribed
from maxicircle DNA provide the edited information. Cell
Burki F, Slachian-Tabrizi K, Minge M, Skjaeveland A, Nikolaev SI,
Jakobsen KS, Pawlowski J. 2007. Phylogenomics reshuffles the
Callahan HA, Litaker RW, Noga EJ. 2002. Molecular taxonomy of the
suborder Bodonina (Order Kinetoplastida), including the
important fish parasite, Ichthyobodo necator. J Eukaryot Micro-
Cavalier-Smith T. 1981. Eukaryote kingdoms: seven or nine? Biosystems
14:461–481.
Delsuc F, Brinkmann H, Philippe H. 2005. Phylogenomics and the
Dolezal D, Jirku M, Maslov DA, Lukes J. 2000. Phylogeny of the
bodonid flagellates (Kinetoplastida) based on small-subunit
endosymbionts of Neoparamoeba spp., relatives of the kineto-
Comparative genomics of trypanosomatid parasitic protozoa.
Etheridge RD, Aphiszheva I, Gershon PD, Aphasizhev R. 2008. 3’
adenylation determines mRNA abundance and monitors completion of RNA editing in T. brucei mitochondria. EMBO J.
27:1596–1608.
Hampl V, Hugo L, Leigh JW, Dacks JB, Lang BF, Simpson AG, Roger AJ.
2009. Phylogenomic analyses support the monophyly of Excavata and resolve relationships among eukaryotic “super-
Kinetoplastid guide RNA biogenesis is dependent on subunits
of the mitochondrial RNA binding complex I and mitochondrial
TbRGG1, an essential protein involved in kinetoplastid RNA
metabolism that is associated with a novel multiprotein complex. RNA 14:970–980.
Jackson AP, Quail MA, Berriman M. 2008. Insights into the genome
sequence of a free-living Kinetoplastid: Bodo saltans (Kineto-
Keeling P, Burger G, Durnford DG, Lang BF, Lee RW, Pearlman RE,
Evol. 20:670–676.
Lartillot N, Philippe H. 2004. A Bayesian mixture model for across-
site heterogeneities in the amino-acid replacement process. Mol
Le SQ, Gascuel O. 2008. An improved general amino acid
Lukès J, Arts GJ, van den Burg J, de Haan A, Opperdooes F, Sloof P,
transcripts of the cryptobiid Trypanoplasma borelli. Embo J.
Marande W, Burger G. 2007. Mitochondrial DNA as a genomic
Maslov DA, Simpson L. 1994. RNA editing and mitochondrial
genomic organization in the cryptobiid kinetoplastid protozoan
kinetoplastid phylogeny using environmental sequences and a
closer outgroup: proposal for a new classification of the class
Osato D, Rogers K, Guo Q, Li F, Richmond G, Klug F, Simpson L.
2009. Uridine insertion/deletion RNA editing in trypanosomatid
39:37–52.
Comparative genomic analysis of three kinetoplastids: in search of the editosome.
Monophyly of the eukaryotic supergroups”.
Philippe H, Lang BF. 2007. Toward resolving the eukaryotic tree:
Philippe H. 2007. Detecting and overcoming systematic errors in
17:1420–1425.
Rodríguez-Ezpeleta N, Brinkmann H, Burey SC, Roure B, Burger G,
of primary photosynthetic eukaryotes: green plants, red algae,
Rodríguez-Ezpeleta N, Brinkmann H, Burger G, Roger AJ, Gray MW,
Philippe H, Lang BF. 2007. Toward resolving the eukaryotic tree:
the phylogenetic positions of jakobids and cercozoans. Curr Biol.
17:1420–1425.
Rodríguez-Ezpeleta N, Brinkmann H, Roure B, Lartillot N, Lang BF,
Philippe H. 2007. Detecting and overcoming systematic errors in


