# 18S rDNA Sequence-Structure Phylogeny of the Trypanosomatida (Kinetoplastea, Euglenozoa) with Special Reference on *Trypanosoma*

**Alyssa R. Borges[1,§], Markus Engstler[1] & Matthias Wolf[2,§*]**

[1]*Department of Cell and Developmental Biology, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany*

[2]*Department of Bioinformatics, Biocenter, University of Würzburg, Am Hubland, 97074 Würzburg, Germany*

[§] *These authors contributed equally to this work*

* *Corresponding author: matthias.wolf@biozentrum.uni-wuerzburg.de (M. Wolf)*

## Abstract

**Background:** Parasites of the order Trypanosomatida are known due to their medical relevance. Trypanosomes cause African sleeping sickness and Chagas disease in South America, and *Leishmania* ROSS, 1903 species mutilate and kill hundreds of thousands of people each year. However, human pathogens are very few when compared to the great diversity of trypanosomatids. Despite the progresses made in the past decades on understanding the evolution of this group of organisms, there are still many open questions which require robust phylogenetic markers to increase the resolution of trees.

**Methods:** Using two known 18S rDNA template structures (from *Trypanosoma cruzi* CHAGAS, 1909 and *Trypanosoma brucei* PLIMMER & BRADFORD, 1899), individual 18S rDNA secondary structures were predicted by homology modeling. Sequences and their secondary structures, automatically encoded by a 12-letter alphabet (each nucleotide with its three structural states, paired left, paired right, unpaired), were simultaneously aligned. Sequence-structure trees were generated by neighbor joining and/or maximum likelihood.

**Results:** With a few exceptions, all nodes within a sequence-structure maximum likelihood tree of 43 representative 18S rDNA sequence-structure pairs are robustly supported (bootstrap support >75). Even a quick and easy sequence-structure neighbor-joining analysis yields accurate results and enables reconstruction and discussion of the big picture for all 240 18S rDNA sequence-structure pairs of trypanosomatids that are currently available.

**Conclusions:** We reconstructed the phylogeny of a comprehensive sampling of trypanosomes evaluated in the context of trypanosomatid diversity, demonstrating that the simultaneous use of 18S rDNA sequence and secondary structure data can reconstruct robust phylogenetic trees.

**Keywords:** Evolution, SSU, Phylogeny, RNA, Secondary Structure, Sequence-Structure, *Trypanosoma*.

## Background

Kinetoplastids (Kinetoplastea HONIGBERG, 1963) are a remarkable group of unicellular organisms. They include free-living and parasite protists of invertebrates, vertebrates, and plants [1–3]. Among them, we find the obligatory parasites of the order Trypanosomatida KENT, 1880, emend VICKERMAN in Moreira et al. 2004 [3], including the human pathogens *T. brucei*, which causes African sleeping sickness, *T. cruzi*, the causative agent of Chagas disease in South America, and *Leishmania* species which infect and harm hundreds of

thousands of people each year [2,4,5]. African trypanosomes are likely the most well-known trypanosomatids. Due to their dixenic life cycle and the extracellular lifestyle in the vertebrate blood, they have evolved interesting, and sometimes unusual, mechanisms to deal with such different environments and challenges imposed by the immune system. Thus, it does not come as a surprise that major discoveries in cell biology have been made in trypanosomes, such as antigenic variation [6], glycolysis compartmented in unique organelles [7], GPI-anchoring of membrane proteins [8], and unprecedented nucleotide modifications [9].

For harboring such a diverse group of organisms, it is unsurprising that the evolution of parasitism inside Kinetoplastea has been intriguing scientists for decades. Given that each parasitic group has closely affiliated free-living relatives and reversion to a free-living state did not occur, it is probable that at least four independent adoptions of obligate parasitism or commensalism have occurred [2,5]. Currently, the earliest diverging lineage inside Trypanosomatida is the genus *Paratrypanosoma* VOTYPKA & LUKES, 2013, represented by one species found in mosquitoes, *Paratrypanosoma confusum* VOTYPKA & LUKES, 2013 [10,11]. The vast majority of trypanosomatids are monoxenic parasites of insects with few dixenic genera due to the capacity of infect vertebrates, such as *Leishmania* and *Trypanosoma* GRUBY, 1843 [2]. Thus, the most likely origin of *Leishmania* and *Trypanosoma* is from within monoxenic trypanosomatids, implicating that their origins were no earlier than 370 million years ago, when the invasion of land by vertebrates occurred [5,12,13]. The transmission of an insect-living trypanosomatid into a warm-blooded host has most likely occurred many times with rare successful cases [5,10,13]. So far, only *Trypanosoma* and *Leishmania* have left surviving descendants and only trypanosomes have adopted an extracellular lifestyle in vertebrates.

Once having passed the vertebrate colonization bottleneck, *Trypanosoma* radiation and adaptation to diverse vertebrate species became an unprecedented evolutionary success story. Today, these parasites prosper in essentially all vertebrate Classes, from fish to birds [13–15].

The 18S rDNA marker has been extensively used to analyze the phylogenetic relationship inside this group [5,12,16–19]. The incorporation of glycosomal glyceraldehyde phosphate dehydrogenase (gGAPDH) to the analysis allowed the generation of more resolved trees, consolidating, for example, the long-lasting question about the monophyly inside *Trypanosoma* [13,15,17–21].

The advent of modern sequencing technologies has greatly advanced our understanding of trypanosomatid phylogeny, with more new genera described in the last decade than within the past century [2,3,5]. These days, trypanosomatid phylogeny has sufficiently advanced to provide a solid framework for comparative studies, with genomic data available for more than just the medically relevant kinetoplastids. Interestingly, the basic layout of trypanosomatid genomes appears to be strikingly similar, with high overall synteny, within and between monoxenic and dixenic species [2]. The constantly growing genome data might become a powerful tool for evolutionary inference. In the future, trypanosomatids will be studied not only as infective agents of devastating neglected tropical diseases, or powerful genetic and cellular model systems, but also to unravel basic principles of the evolution of unicellular eukaryotes. What we know today is just the tip of an iceberg. The origin of *Trypanosoma*, for example, remains enigmatic. Further, the presence of prokaryotic endosymbionts and viruses in trypanosomatids, or the full biodiversity and ecological role of insect trypanosomatids remain superficially explored [2,22].

Here, we present the first large scale study in which trypanosomatid RNA secondary structure is used as an additional source of phylogenetic information. We use 18S rRNA sequence-structure data simultaneously in inferring alignments and trees. This approach was recently reviewed and shown to increase robustness and accuracy of reconstructed phylogenies [23,24]. So far, all conclusions have been made with multigene trees. The most important result of our study is that there are only a few places in our robustly supported trees where branching does not match with multigene phylogenomic trees. It is these discrepancies with

which we focus our discussion, as they are potentially critical branches that are ambiguous and require more attention.

# Methods

## Taxon Sampling, Secondary Structure Prediction, Sequence–Structure Alignment, and Phylogenetic Tree Reconstruction

We have used all 240 SSU 18S ribosomal RNA gene sequences from Trypanosomatida available at NCBI (GenBank) with a sequence length >1500 nucleotides and with a full taxonomic lineage down to a complete species name. Secondary structures of the 18S rDNA sequences were obtained by homology modeling [25] using the ITS2 database [26] as well as *T. cruzi* (AF245382) and *T. brucei* (M12676) as templates (Fig S1). The two template-secondary structures (without pseudoknots) were obtained from the Comparative RNA Web (CRW) site [27]. For sequence-structure alignments, the four nucleotides multiplied by three states (unpaired, paired left and paired right) are encoded by a 12-letter alphabet [24]. Using a specific 12×12 sequence-structure scoring matrix [28], global multiple sequence-structure alignments were automatically generated in ClustalW2 1.83 [29] as implemented in 4SALE 1.7.1 [28,30]. Based on Keller et al. [23], using 12-letter encoded sequences, sequence-structure neighbor-joining (NJ) trees were determined using ProfDistS [31]. Further, using 12-letter encoded sequences, a sequence-structure maximum likelihood (ML) tree [32] - for only a representative subset of 43 sequence-structure pairs - was calculated using phangorn [33] as implemented in the statistical framework R [34]. The R script is available from the 4SALE homepage at http://4sale.bioapps.biozentrum.uni-wuerzburg.de [24]. Bootstrap support for all sequence-structure trees was estimated based on 100 pseudo-replicates. Trees were rooted with non-*Trypanosoma* sequences from Trypanosomatida.

## Results and discussion

From the analysis of 240 18S rDNA sequence-structure pairs (Fig. 1) and selection of 43 different species (Fig. 2), we obtained trees supported by high bootstraps values (> 75) on sister groups displaying the following *Trypanosoma* clades: the *Trypanosoma pestanai* BETTENCOURT & FRANCA, 1905 clade, represented in our tree by this species found in the Eurasian badger [13]; the *T. brucei* clade, consisting of trypanosome species naturally transmitted by tsetse flies, such as *Trypanosoma vivax* ZIEMANN, 1905, *Trypanosoma congolense* BRODEN, 1904, *Trypanosoma godfrey* MCNAMARA ET AL., 1994, *Trypanosoma simiae* BRUCE ET AL., 1913, *Trypanosoma equiperdum* DOFLEIN, 1901, *Trypanosoma evansi* STEEL, 1885, and *T. brucei* [13,17,19,21]; the *T. cruzi* clade, comprising mammalian trypanosomes with worldwide distribution, such as *T. cruzi*, *Trypanosoma rangeli* TEJERA, 1920, and *Trypanosoma wauwau* TEIXEIRA & CAMARGO, 2016, endemic of Latin America, *Trypanosoma conorhini* (DONOVAN, 1909) SHORTT & SWAMINATH, 1928 found in Europe, South America and Africa, and *Trypanosoma dionisii* BETTENCOURT & FRANCA, 1905 distributted in Latin America, Africa, Asia and Europe [19–21,35]; the *Trypanosoma lewisi* (KENT, 1880) LAVERAN & MESNIL, 1901 clade, including the rodent parasites *Trypanosoma microti* LAVERAN & PETTIT, 1910, *Trypanosoma grosi* LAVERAN & PETTIT, 1909 and *T. lewisi* [36]; the Crocodilian clade, harboring *Trypanosoma grayi* NOVY, 1906 from Africa and *Trypanosoma ralphi* TEIXEIRA & CAMARGO, 2013 from South America [15,37]; the Avian clade, with *Trypanosoma corvi* STEPHENS & CHRISTOPHERS, 1908, *Trypanosoma avium* DANILEWSKY, 1885 and *Trypanosoma thomasbancrofti* SLAPETA, 2016 [38]; the *Trypanosoma theileri* LAVERAN, 1902 clade, with *T. theileri*, a worldwide distributted cattle parasite, and the subclade representant *Trypanosoma cyclops* WEINMAN, 1972 [13,21]; and the Aquatic clade, harboring trypanosomes from fish, anurans and platypus [14,39,40]. Interestingly, the lizard/snake clade is also represented in our tree with *Trypanosoma varani* WENYON, 1908, a snake trypanosome, branching together with the mammal parasite *Trypanosoma freitasi* REGO

153 ET AL., 1957. The branching of marsupial and rodent trypanosomes inside this clade has been
154 previously observed [36,41]. Thus, our analysis corroborates the existence of the lizard-
155 snake/marsupial-rodent clade composed by trypanosomes transmitted by sandflies [36].
156 The phylogenetic analyses using sequence-structure data of 18S rDNA (Fig. 2) supports the
157 monophyly of *Trypanosoma* as previously observed in trees constructed with partial/
158 complete sequences of 18S rDNA and/or gGAPDH sequences [15,17–19]. Intriguingly, in the
159 tree obtained using a greater number of sequences (Fig. 1) *Strigomonas culicis* (WALLACE &
160 JOHNSON, 1961) TEIXEIRA & CAMARGO, 2011 (U05679-1 and HQ659564-1) appear as a basal
161 group of African trypanosomes (Fig. 1). To date, studies on Trypanosomatida showed a basal
162 position of *Trypanosoma* in relation to *Strigomonas* LWOFF & LWOFF, 1931 [2,22,42]. Of
163 notice, *S. culicis* is a monoxenic parasite of the order Diptera [22], the same order of the well-
164 known vector of *T. brucei* clade trypanosomes, the tsetse fly. Although it may be tempting to
165 speculate the derivation of the tsetse lifestyle of African trypanosomes from *Strigomonas*, we
166 cannot exclude the possibility of poor sequence assembly, which could interfere with the
167 topology observed. Finally, one sequence of the bat parasite *T. dionisii* clustered within the *T.*
168 *brucei* clade (Fig. 1). This species is distributed worldwide, with its origin in Africa, and
169 presents a high phyletic diversity [35]. However, its branching inside *T. cruzi* clade is strongly
170 supported [13,17–19,21].
171 The first branching of *Trypanosoma* (Fig. 2) forms two major groups: one lineage composed
172 by *T. brucei* and *T. pestanai* clades, and another with trypanosomes from Terrestial (*T. cruzi*,
173 *T. lewisi*, *T. theileri*, snake-lizard/marsupial-rodent, avian and crocodilian clades) and Aquatic
174 lineages. Thus, our tree corroborates the hypothesis of the independent evolutionary history of
175 both human pathogens, *T. brucei* and *T. cruzi* [13]. The topology of our tree shows the
176 Aquatic clade as a solid lineage, in accordance with previous observations [15,18,19,21].
177 However, the origin of this clade is still under debate. Many studies using different DNA
178 markers, such as long (> 1.4 kb) 18S rDNA sequences, v7v8 hypervariable region of 18S

rDNA and/or partial sequences of gGAPDH, showed either an early division between Aquatic and Terrestrial lineages as a single event [15,18,19,21] or in subsequent events with amphibian trypanosomes and *Trypanosoma therezieni* BRYGOO, 1963 at the basis of *Trypanosoma* [17,43,44]. Interestingly, our tree suggests a later evolution of the Aquatic clade from Terrestrial trypanosomes (Fig. 2), which agrees with the insect-first hypothesis [13,36]. This hypothesis assumes that trypanosomes were originated from a monogenetic insect parasite that adapted to live inside terrestrial vertebrates and later spread to leeches and other aquatic animals, most likely through amphibians [13].

Trypanosomes of the *T. brucei* clade are virtually restricted to Africa, having an exception in *T. vivax* [45,46]. The early divergence of *T. vivax* inside the *T. brucei* clade (Fig. 1, 2) is in accordance to previous results showing a higher evolutionary rate of this species among the Salivarian trypanosomes [19,21,47]. It is interesting to consider that a previous analysis of 18S rDNA sequences revealed that members of the *T. brucei* clade show an evolutionary rate higher than other trypanosomes [47]. However, this high divergence has proven not to alter the topology of sequence-based trees [17]. Inside this clade, a low bootstrap value (ML = 53) is observed in the differentiation between *T. brucei* and *T. evansi* (Fig. 2), suggesting an unresolved positioning. In fact, the relationship between these species is controversial, with results supporting either *T. evansi* as a subspecies of *T. brucei* or showing great diversity between both depending on the *T. evansi* strains [48,49].

Regarding the other major group of our analysis (Terrestrial/Aquatic lineages), the first branching inside this group suggests the differentiation of the snake-lizard/marsupial-rodent clade as a basal group of other trypanosomes (Fig. 2). However, other studies have suggested avian trypanosomes as a basal group among terrestrial lineages [13,17]. This can be associated with the low bootstrap values of our tree in either three events: the snake-lizard/marsupial-rodent clade (ML = 54) differentiation, the divergence of crocodilian

trypanosomes (ML = 55), and the internal branch of avian trypanosomes (ML = 61), which will be further explored in our discussion.

The *T. cruzi* and *T. lewisi* clades appear as sister groups in our analysis (Fig. 2), as has been previously demonstrated [17–19,21]. The *T. cruzi* clade can be subdivided into three subclades: Schyzotrypanum, *T. wauwau* (and other Neotropical bat trypanosomes, *Trypanosoma noyesi* BOTERO & COOPER, 2016, and *Trypanosoma livingstonei* TEIXEIRA & CAMARGO, 2013) and *T. rangeli*/*T. conorhini* [21,35]. The specific sequences of *Trypanosoma minasense* CHAGAS, 1908 and *Trypanosoma leeuwenhoeki* SHAW, 1969 grouped with *T. rangeli* in our tree were previously considered synonyms of this species by 18S rDNA sequence analysis, explaining their positioning [13,50]. Regarding the *T. lewisi* clade, our results suggest the existence of two subclades inside the group (ML = 100), one harboring *T. microti,* and the other with *T. lewisi* and *T. grosi*. This finding is in accordance with a recent analysis of long fragments of 18S rDNA which demonstrated this subdivision despite the similarities in the v7v8 hypervariable region [51].

Concerning avian trypanosomes, our tree reflects previous findings in both the divergence between *T. avium* and *T. corvi* and the highly supported (ML = 100) proximity between *T. avium* and *T. thomasbancrofti* [38,50]. A low bootstrap value (ML = 61) is observed in the divergence of *T. corvi* and *T. avium*/ *T. thomasbancrofti*. It is interesting to note that we currently have two topologies known in literature, with the possibility of paraphyly demonstrated by analysis of long sequences of 18S rDNA [19,38]. This, however, was not observed in trees constructed with v7v8 hypervariable region of 18S rDNA, gGAPDH sequences, or concatenated trees using both [15,37,52]. Thus, our tree indicates the need for a better resolution on avian trypanosome positioning. Considering that we used only three species of avian and two species of crocodilian trypanosomes in our reconstruction, our approach represents an interesting method to be applied in further studies.

In our analysis we see the crocodilian/alligator trypanosomes (*T. grayi* and *T. ralphi*) branching together with a high support value (ML = 100) (Fig. 2). Although *T. grayi* is found in Africa and *T. ralphi* in South America [15,37,52], in a tree with distant external groups like ours, this topology is expected due to their proximity inside the Crocodilian clade [15,37]. Our tree reflects the proximity of the crocodilian trypanosomes with *T. cruzi* clade, as previously observed through full genome analysis [53]. Interestingly, crocodilian trypanosomes, such as *T. grayi* and *Trypanosoma kaiowa* TEIXEIRA & CAMARGO, 2019 are tsetse-transmitted species that are not restricted to the sub-Saharan belt [15,37,53], suggesting higher adaptive plasticity of crocodilian trypanosomes.

The trypanosomes of the Aquatic lineage branched together (Fig. 2). The subgroups observed are anuran trypanosomes (*Trypanosoma rotatorium* (MAYER, 1843) LAVERAN, 1901, *Trypanosoma mega* DUTTON & TODD, 1903, *Trypanosoma fallisi* MARTIN & DESSER, 1990, *Trypanosoma ranarum* (LANKESTER, 1871) DANILEWSKY, 1885, and *Trypanosoma neveulemairei* BRUMPT, 1928) and fish trypanosomes (*Trypanosoma siniperca* CHANG, 1964, *Trypanosoma ophiocephali* CHEN, 1964, *Trypanosoma cobitis* MITROPHANOW, 1884, *Trypanosoma granulosum* LAVERAN & MESNIL, 1902, *Trypanosoma pleuronectidium* ROBERTSON, 1906) along with the platypus parasite *Trypanosoma binneyi* MACKERRAS, 1959, which is in accordance to the literature [14,15,39,40]. Interestingly, the anuran parasite, *Trypanosoma chattoni* MATHIS & LEGER, 1911, appears in our analysis more related to fish and platypus trypanosomes than to the anuran clade. This positioning of *T. chattoni* was shown in a previous study using complete 18S rDNA sequences and non-trypanosomes as the outgroup [54]. However, recent trees using complete 18S rDNA sequences and concatenated analysis of v7v8 hypervariable region and gGAPDH rooted by other trypanosomes sustained a monophyletic anuran clade [39,40]. Thus, *T. chattoni* positioning in our tree can be related to the use of non-trypanosomes as the outgroup.

## Conclusions

18S rDNA is one of the most used markers in inferring phylogenies at higher taxonomic levels. However, this molecule has often been claimed to be inadequate for reconstructing phylogenetic relationships at lower taxonomic levels, in particular, because of its conservative rate of evolution. Thus, the use of the v7v8 hypervariable region has become popular inside the trypanosome community as a barcode suitable to describe new species along with gGAPDH [13]. However, some questions regarding the positioning of the groups still need to be answered. To this, evaluating bigger sequences and using different genetic markers can bring more information to the analysis improving resolution of the trees [13]. In this work, we demonstrate that the simultaneous use of 18S rDNA sequence and secondary structure data (i.e., the consideration of the individual secondary structures of the rRNA genes) could indeed provide important information for reconstructing more robust phylogenetic trees. Our topology highlights the need for further exploration of some groups, such as the avian and snake-lizard/marsupial-rodent clades, which are less explored in trypanosome phylogenies. Therefore, this study provides an additional fundament for upcoming phylogenetic reconstructions and/or barcoding approaches which can be used not only in trypanosomatids.

## Declarations

### Ethics approval and consent to participate

Not applicable

### Consent for publication

Not applicable

## Availability of data and materials

The sequence data analysed in this paper is publicly available in NCBI (GenBank) and their respective accession numbers are included in the trees. Data supporting the conclusions of this article are included within the article. Data and materials are available upon reasonable request to the corresponding author.

## Competing interests

The authors declare that they have no competing interests.

## Funding

Ph.D. scholarship was granted to Alyssa Rossi Borges by Brazilian agency CAPES (program: CAPES/DAAD - Call No. 22/2018; process 88881.199683/2018-01). ME is supported by DFGgrants EN-305.

## Authors' contributions

MW conceived the study and performed the bioinformatic analysis. ARB analyzed the data and majorly contributed to the manuscript writing. ME and MW assisted with data analysis and manuscript writing. All authors read and approved the final manuscript.

## Acknowledgments

We thank Julius Lukeš (Institute of Parasitology, Biology Centre CAS, Czech Republic) for fruitful discussions and a pre-review of our manuscript, Jaime Lisack for proof-reading, and Elisabeth Meyer-Natus for obtaining the electron microscopy image used in the graphical abstract.

# References

306 1. Cavalier-Smith T. Higher classification and phylogeny of Euglenozoa. Eur J Protistol.
307 2016;56:250–76.

308 2. Lukeš J, Butenko A, Hashimi H, Maslov DA, Votýpka J, Yurchenko V. Trypanosomatids
309 are much more than just trypanosomes: clues from the expanded family tree. Trends Parasitol.
310 2018;34:466–80.

311 3. Adl SM, Bass D, Lane CE, Lukeš J, Schoch CL, Smirnov A, et al. Revisions to the
312 classification, nomenclature, and diversity of eukaryotes. J Euk Microbiol. 2019;66:4–119.

313 4. WHO. Research priorities for Chagas disease, human African trypanosomiasis and
314 leishmaniasis. WHO; 2012.
315 http://www.who.int/trypanosomiasis_african/resources/who_trs_975/en/

316 5. Lukeš J, Skalický T, Týč J, Votýpka J, Yurchenko V. Evolution of parasitism in
317 kinetoplastid flagellates. Mol Biochem Parasitol. 2014;195:115–22.

318 6. Vickerman K. Antigenic variation in trypanosomes. Nature. 1978;273:613–7.

319 7. Hart DT, Baudhuin P, Opperdoes FR, de Duve C. Biogenesis of the glycosome in
320 *Trypanosoma brucei*: the synthesis, translocation and turnover of glycosomal polypeptides.
321 EMBO J. 1987;6:1403–11.

322 8. Menon AK, Mayor S, Ferguson MA, Duszenko M, Cross GA. Candidate
323 glycophospholipid precursor for the glycosylphosphatidylinositol membrane anchor of
324 *Trypanosoma brucei* variant surface glycoproteins. J Biol Chem. 1988;263:1970–7.

325 9. Gommers-Ampt JH, Van Leeuwen F, de Beer AL, Vliegenthart JF, Dizdaroglu M,
326 Kowalak JA, et al. beta-D-glucosyl-hydroxymethyluracil: a novel modified base present in the
327 DNA of the parasitic protozoan *T. brucei*. Cell. 1993;75:1129–36.

328 10. Flegontov P, Votýpka J, Skalický T, Logacheva MD, Penin AA, Tanifuji G, et al.
329 *Paratrypanosoma* is a novel early-branching trypanosomatid. Curr Biol. 2013;23:1787–93.

330 11. Skalický T, Dobáková E, Wheeler RJ, Tesařová M, Flegontov P, Jirsová D, et al.
331 Extensive flagellar remodeling during the complex life cycle of *Paratrypanosoma*, an early-
332 branching trypanosomatid. Proc Natl Acad Sci USA. 2017;114:11757–62.

333 12. Stevens JR, Gibson W, Stevens JR, Gibson W. The molecular evolution of trypanosomes.
334 Parasitol Today. 1999;15:432–7.

335 13. Hamilton PB, Stevens JR. Classification and phylogeny of *Trypanosoma cruzi*. In:
336 Telleria J, Tibayrenc M, editors. American Trypanosomiasis Chagas Disease. 2nd ed.
337 London: Elsevier; 2017. p. 321–44.

338 14. Lemos M, Fermino BR, Simas-Rodrigues C, Hoffmann L, Silva R, Camargo EP, et al.
339 Phylogenetic and morphological characterization of trypanosomes from Brazilian armoured
340 catfishes and leeches reveal high species diversity, mixed infections and a new fish
341 trypanosome species. Parasit Vectors. 2015;8:573.

342 15. Fermino BR, Paiva F, Viola LB, Rodrigues CMF, Garcia HA, Campaner M, et al. Shared
343 species of crocodilian trypanosomes carried by tabanid flies in Africa and South America,

including the description of a new species from caimans, *Trypanosoma kaiowa* n. sp. Parasit Vectors. 2019;12:225.

16. Maslov DA, Podlipaev SA, Lukes J. Phylogeny of the kinetoplastida: taxonomic problems and insights into the evolution of parasitism. Mem Inst Oswaldo Cruz. 2001;96:397–402.

17. Hamilton PB, Stevens JR, Gaunt MW, Gidley J, Gibson WC. Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. Int J Parasitol. 2004;34:1393–404.

18. Hamilton PB, Stevens JR, Gidley J, Holz P, Gibson WC. A new lineage of trypanosomes from Australian vertebrates and terrestrial bloodsucking leeches (Haemadipsidae). Int J Parasitol. 2005;35:431–43.

19. Hamilton PB, Gibson WC, Stevens JR. Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. Mol Phylogenet Evol. 2007;44:15–25.

20. Hamilton PB, Adams ER, Njiokou F, Gibson WC, Cuny G, Herder S. Phylogenetic analysis reveals the presence of the *Trypanosoma cruzi* clade in African terrestrial mammals. Infect Genet Evol. 2009;9:81–6.

21. Lima L, Espinosa-Álvarez O, Pinto CM, Cavazzana M, Pavan AC, Carranza JC, et al. New insights into the evolution of the *Trypanosoma cruzi* clade provided by a new trypanosome species tightly linked to Neotropical *Pteronotus* bats and related to an Australian lineage of trypanosomes. Parasit Vectors. 2015;8:657.

22. Teixeira MMG, Borghesan TC, Ferreira RC, Santos MA, Takata CSA, Campaner M, et al. Phylogenetic validation of the genera *Angomonas* and *Strigomonas* of trypanosomatids harboring bacterial endosymbionts with the description of new species of trypanosomatids and of proteobacterial symbionts. Protist. 2011;162:503–24.

23. Keller A, Förster F, Müller T, Dandekar T, Schultz J, Wolf M. Including RNA secondary structures improves accuracy and robustness in reconstruction of phylogenetic trees. Biol Direct. 2010;5:4.

24. Wolf M, Koetschan C, Müller T. ITS2, 18S, 16S or any other RNA - simply aligning sequences and their individual secondary structures simultaneously by an automatic approach. Gene. 2014;546:145–9.

25. Hall BS, Smith E, Langer W, Jacobs LA, Goulding D, Field MC. Developmental variation in Rab11-dependent trafficking in *Trypanosoma brucei*. Eukaryot Cell. 2005;4:971–80.

26. Ankenbrand MJ, Keller A, Wolf M, Schultz J, Förster F. ITS2 Database V: Twice as Much. Mol Biol Evol. 2015;32:3030–2.

27. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, et al. The Comparative RNA Web (CRW) Site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics. 2002;3:2.

28. Seibel PN, Müller T, Dandekar T, Schultz J, Wolf M. 4SALE – A tool for synchronous RNA sequence and secondary structure alignment and editing. BMC Bioinformatics. 2006;7:498.

29. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, et al. Clustal W and Clustal X version 2.0. Bioinformatics. 2007;23:2947–8.

30. Seibel PN, Müller T, Dandekar T, Wolf M. Synchronous visual analysis and editing of RNA sequence and secondary structure alignments using 4SALE. BMC Res Notes. 2008;1:91.

31. Wolf M, Ruderisch B, Dandekar T, Schultz J, Müller T. ProfDistS: (profile-) distance based phylogeny on sequence--structure alignments. Bioinformatics. 2008;24:2401–2.

32. Felsenstein J. Evolutionary trees from gene frequencies and quantitative characters: finding maximum likelihood estimates. Evolution. 1981;35:1229–42.

33. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27:592–3.

34. R Core Team. R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna.; 2014.

35. Clément L, Dietrich M, Markotter W, Fasel NJ, Monadjem A, López-Baucells A, et al. Out of Africa: The origins of the protozoan blood parasites of the *Trypanosoma cruzi* clade found in bats from Africa. Mol Phylogenet Evol. 2020;145:106705.

36. Ortiz PA, Garcia HA, Lima L, da Silva FM, Campaner M, Pereira CL, et al. Diagnosis and genetic analysis of the worldwide distributed *Rattus*-borne *Trypanosoma* (*Herpetosoma*) *lewisi* and its allied species in blood and fleas of rodents. Infect Genet Evol. 2018;63:380–90.

37. Fermino BR, Viola LB, Paiva F, Garcia HA, de Paula CD, Botero-Arias R, et al. The phylogeography of trypanosomes from South American alligatorids and African crocodilids is consistent with the geological history of South American river basins and the transoceanic dispersal of *Crocodylus* at the Miocene. Parasit Vectors. 2013;6:313.

38. Šlapeta J, Morin-Adeline V, Thompson P, McDonell D, Shiels M, Gilchrist K, et al. Intercontinental distribution of a new trypanosome species from Australian endemic Regent Honeyeater (*Anthochaera phrygia*). Parasitology. 2016;143:1012–25.

39. Attias M, Sato LH, Ferreira RC, Takata CSA, Campaner M, Camargo EP, et al. Developmental and ultrastructural characterization and phylogenetic analysis of *Trypanosoma herthameyeri* n. sp. of Brazilian Leptodactilydae frogs. J Euk Microbiol. 2016;63:610–22.

40. Spodareva VV, Grybchuk-Ieremenko A, Losev A, Votýpka J, Lukeš J, Yurchenko V, et al. Diversity and evolution of anuran trypanosomes: insights from the study of European species. Parasit Vectors. 2018;11:447.

41. Dobigny G, Poirier P, Hima K, Cabaret O, Gauthier P, Tatard C, et al. Molecular survey of rodent-borne *Trypanosoma* in Niger with special emphasis on *T. lewisi* imported by invasive black rats. Acta Trop. 2011;117:183–8.

42. Du Y, Maslov DA, Chang K-P. Monophyletic origin of (division proteobacterial endosymbionts and their coevolution with insect trypanosomatid protozoa *Blastocrithidia culicis* and *Crithidia* spp. Proc Natl Acad Sci USA. 1994;91:8437–8441.

43. Botero A, Thompson CK, Peacock CS, Clode PL, Nicholls PK, Wayne AF, et al. Trypanosomes genetic diversity, polyparasitism and the population decline of the critically

endangered Australian marsupial, the brush tailed bettong or woylie (*Bettongia penicillata*). Int J Parasitol Parasites Wildl. 2013;2:77–89.

44. Botero A, Cooper C, Thompson CK, Clode PL, Rose K, Thompson RCA. Morphological and phylogenetic description of *Trypanosoma noyesi* sp. nov.: an Australian wildlife trypanosome within the *T. cruzi* clade. Protist. 2016;167:425–39.

45. Rodrigues AC, Neves L, Garcia HA, Viola LB, Marcili A, Da Silva FM, et al. Phylogenetic analysis of *Trypanosoma vivax* supports the separation of South American/West African from East African isolates and a new *T. vivax*-like genotype infecting a nyala antelope from Mozambique. Parasitology. 2008;135:1317–28.

46. Silva Pereira S, de Almeida Castilho Neto KJG, Duffy CW, Richards P, Noyes H, Ogugo M, et al. Variant antigen diversity in *Trypanosoma vivax* is not driven by recombination. Nat Commun. 2020;11:844.

47. Stevens J, Rambaut A. Evolutionary rate differences in trypanosomes. Infect Genet Evol. 2001;1:143–50.

48. Carnes J, Anupama A, Balmer O, Jackson A, Lewis M, Brown R, et al. Genome and phylogenetic analyses of *Trypanosoma evansi* reveal extensive similarity to *T. brucei* and multiple independent origins for dyskinetoplasty. PLoS Negl Trop Dis. 2015;9.

49. Kamidi CM, Saarman NP, Dion K, Mireji PO, Ouma C, Murilla G, et al. Multiple evolutionary origins of *Trypanosoma evansi* in Kenya. PLoS Negl Trop Dis. 2017;11.

50. Sato H, Leo N, Katakai Y, Takano J, Akari H, Nakamura S, et al. Prevalence and molecular phylogenetic characterization of *Trypanosoma* (*Megatrypanum*) *minasense* in the peripheral blood of small neotropical primates after a quarantine period. J Parasitol. 2008;94:1128–38.

51. Egan SL, Taylor CL, Austen JM, Banks PB, Ahlstrom LA, Ryan UM, et al. Molecular identification of the *Trypanosoma* (*Herpetosoma*) *lewisi* clade in black rats (*Rattus rattus*) from Australia. Parasitol Res. 2020;119:1691–6.

52. Fermino BR, Paiva F, Soares P, Tavares LER, Viola LB, Ferreira RC, et al. Field and experimental evidence of a new caiman trypanosome species closely phylogenetically related to fish trypanosomes and transmitted by leeches. Int J Parasitol Parasites Wildl. 2015;4:368–78.

53. Kelly S, Ivens A, Manna PT, Gibson W, Field MC. A draft genome for the African crocodilian trypanosome *Trypanosoma grayi*. Sci Data. 2014;1:140024.

54. Martin DS, Wright A-DG, Barta JR, Desser SS. Phylogenetic position of the giant anuran trypanosomes *Trypanosoma chattoni*, *Trypanosoma fallisi*, *Trypanosoma mega*, *Trypanosoma neveulemairei*, and *Trypanosoma ranarum* inferred from 18S rRNA gene sequences. J Parasitol. 2002;88:566–71.

# Figure legends:

**Fig. 1: 18S rDNA sequence-structure neighbor-joining (NJ) tree** obtained by ProfDistS [31]. All 240 18S rDNA sequences from Trypanosomatida (Kinetoplastea, Euglenozoa) available at NCBI (GenBank) with a sequence length >1500 nucleotides and with a full taxonomic lineage down to a complete species name have been used for the analysis. For tree reconstruction the global multiple sequence-structure alignment (.xfasta format) as derived by 4SALE [28,30] was automatically encoded by a 12-letter alphabet [24]. GenBank accession numbers accompany each taxon name. Key taxa are off and on marked in gray and additionally named alongside the tree. Non-monophyletic taxa are indicated by quotation marks. Singletons are highlighted red and polyphyletic taxa are highlighted blue. The scale bar indicates evolutionary distances. The tree is rooted at non-*Trypanosoma* sequences.

**Fig. 2: 18S rDNA sequence-structure maximum likelihood (ML) tree**, representative subset of 43 sequences from Fig. 1, obtained by phangorn as implemented in R [33]. Bootstrap values from 100 pseudo-replicates, mapped at the appropriate internodes, are from maximum likelihood- (ML) and neighbor-joining- (NJ, obtained by ProfDistS, Wolf et al. [31]) analyses. For NJ tree reconstruction the global multiple sequence-structure alignment (.xfasta format) as derived by 4SALE [28,30] was automatically encoded by a 12-letter alphabet [24]. For ML tree reconstruction the “one letter encoded” fasta format (12-letter alphabet) as derived by 4SALE [28,30] was used. GenBank identifiers accompany each taxon name. The scale bar indicates evolutionary distances. Highly supported branches are indicated by thicker lines. The tree is rooted at *Crithridia mellificae* (KM980182) and *Leishmania amazonensis* (JX030087). Clades discussed in the text are highlighted.

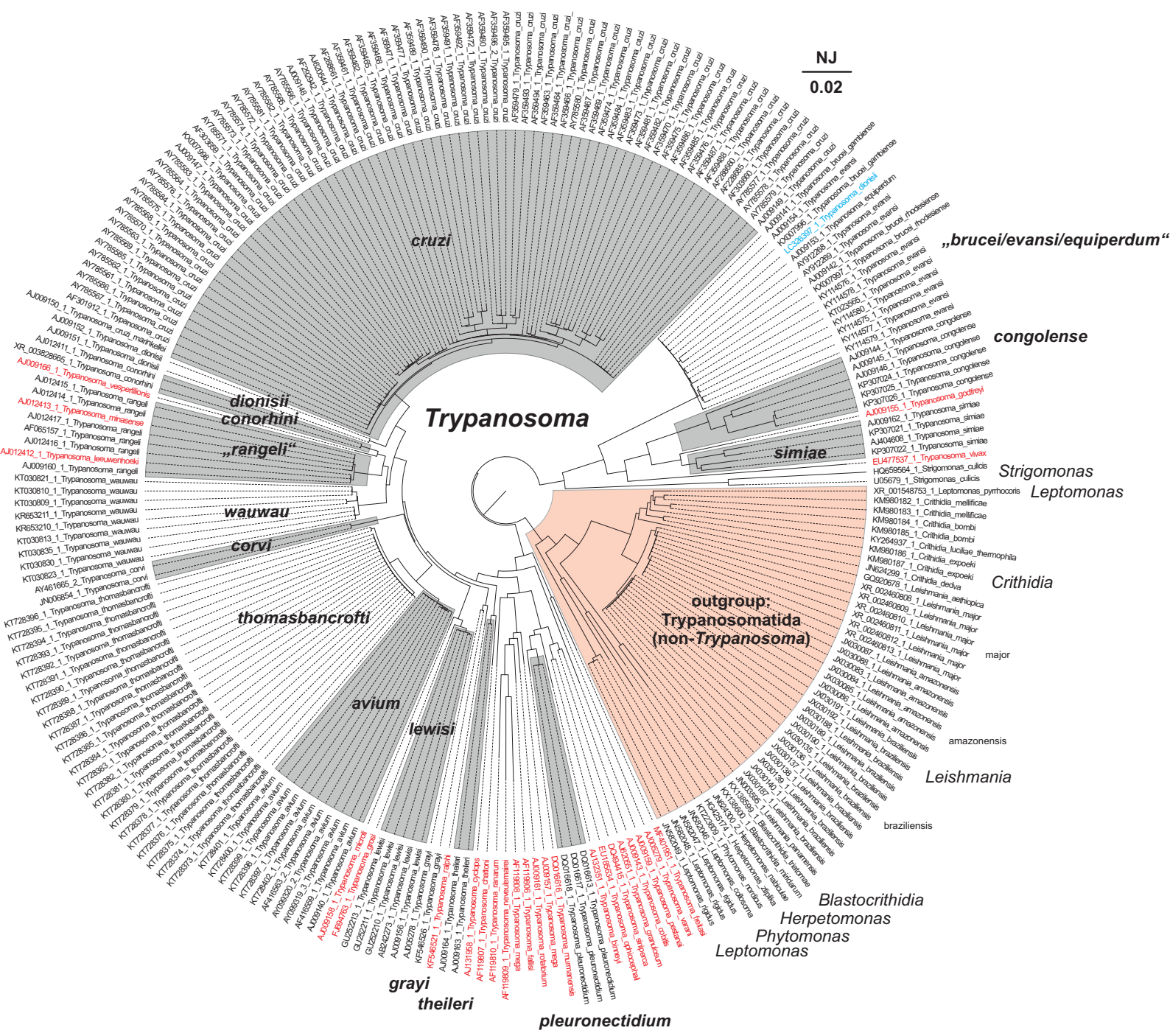
# Supplementary Material

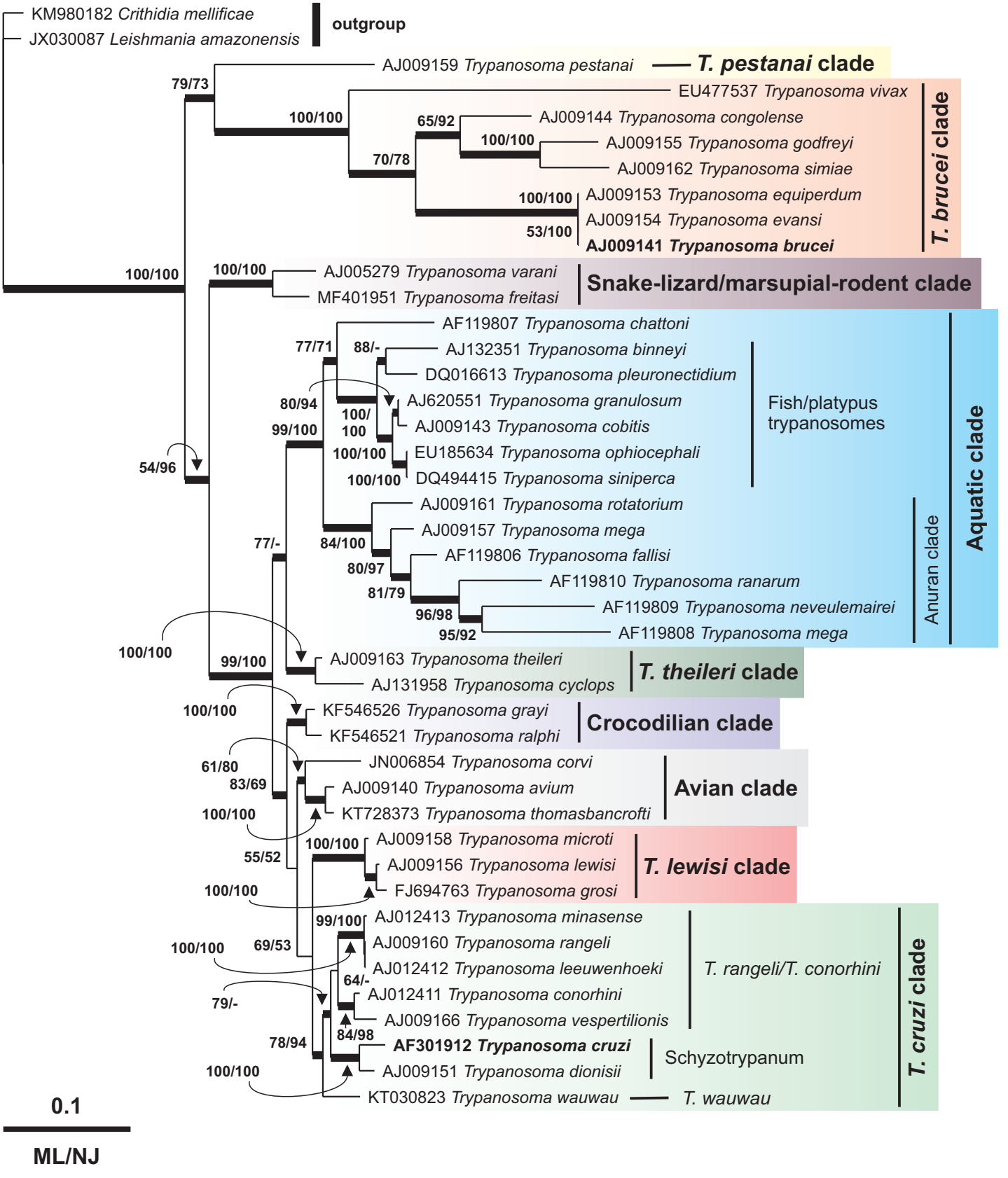
487

488 **Fig. S1:** *Trypanosoma brucei* (M12676). Secondary structure: small subunit ribosomal RNA

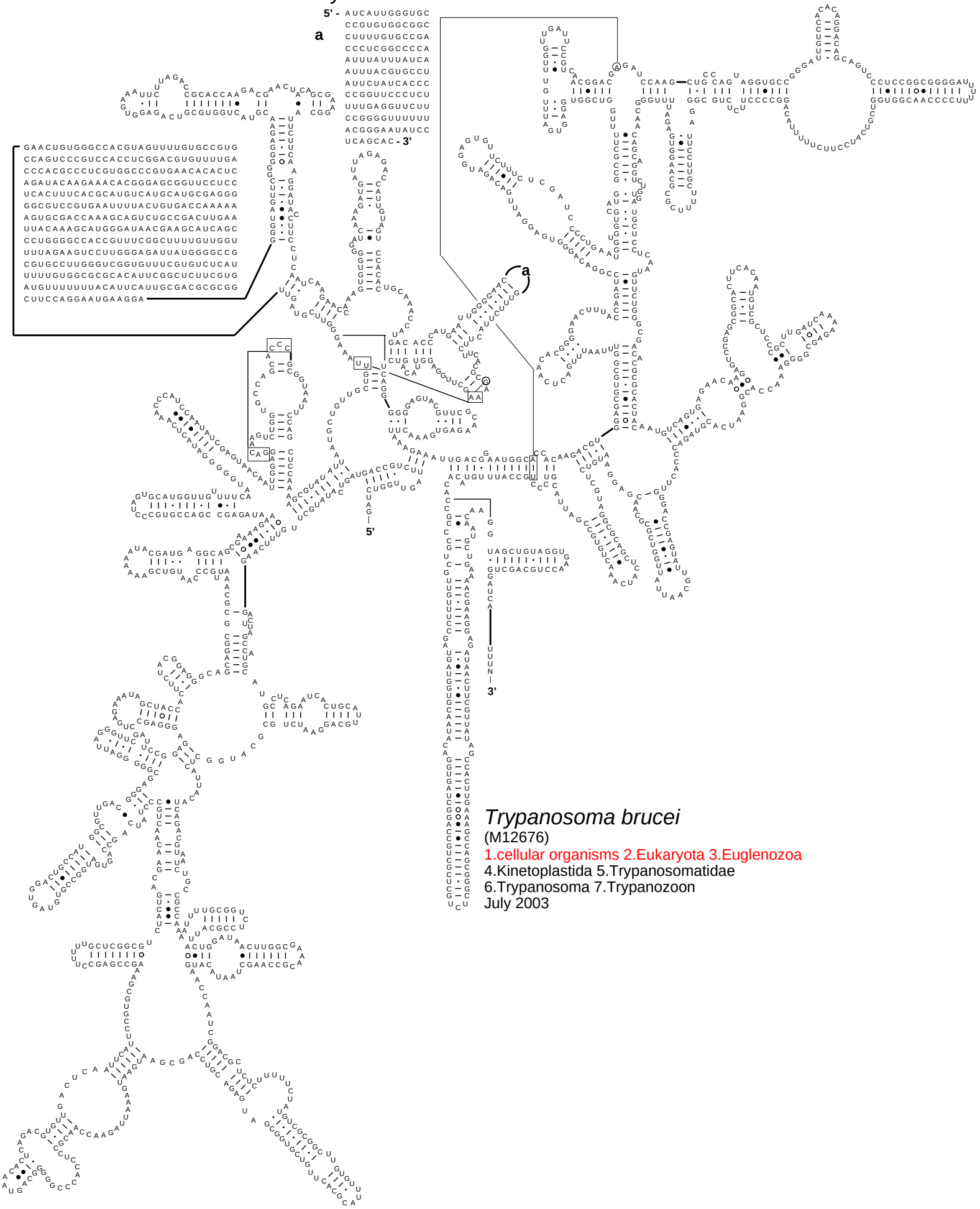
489 taken from the Comparative RNA Web (CRW) site [27].

490


NJ
0.02
Trypanosoma
cruzi
„brucei/evansi/equiperdum“
congolense
simiae
dionisii
conorhini
„rangeli“
wauwau
corvi
thomasbancrofti
avium
lewisi
grayi
theileri
pleuronectidium
outgroup:
Trypanosomatida
(non-Trypanosoma)
Strigomonas
Leptomonas
Crithidia
major
amazonensis
braziliensis
Leishmania
Blastocrithidia
Herpetomonas
Phytomonas
Leptomonas


KM980182 Crithidia mellificae
JX030087 Leishmania amazonensis
outgroup
AJ009159 Trypanosoma pestanai
T. pestanai clade
EU477537 Trypanosoma vivax
AJ009144 Trypanosoma congolense
AJ009155 Trypanosoma godfreyi
AJ009162 Trypanosoma simiae
AJ009153 Trypanosoma equiperdum
AJ009154 Trypanosoma evansi
AJ009141 Trypanosoma brucei
T. brucei clade
AJ005279 Trypanosoma varani
MF401951 Trypanosoma freitasi
Snake-lizard/marsupial-rodent clade
AF119807 Trypanosoma chattoni
AJ132351 Trypanosoma binneyi
DQ016613 Trypanosoma pleuronectidium
AJ620551 Trypanosoma granulosum
AJ009143 Trypanosoma cobitis
EU185634 Trypanosoma ophiocephali
DQ494415 Trypanosoma siniperca
Fish/platypus trypanosomes
AJ009161 Trypanosoma rotatorium
AJ009157 Trypanosoma mega
AF119806 Trypanosoma fallisi
AF119810 Trypanosoma ranarum
AF119809 Trypanosoma neveulemairei
AF119808 Trypanosoma mega
Anuran clade
Aquatic clade
AJ009163 Trypanosoma theileri
AJ131958 Trypanosoma cyclops
T. theileri clade
KF546526 Trypanosoma grayi
KF546521 Trypanosoma ralphi
Crocodilian clade
JN006854 Trypanosoma corvi
AJ009140 Trypanosoma avium
KT728373 Trypanosoma thomasbancrofti
Avian clade
AJ009158 Trypanosoma microti
AJ009156 Trypanosoma lewisi
FJ694763 Trypanosoma grosi
T. lewisi clade
AJ012413 Trypanosoma minasense
AJ009160 Trypanosoma rangeli
AJ012412 Trypanosoma leeuwenhoeki
AJ012411 Trypanosoma conorhini
AJ009166 Trypanosoma vespertilionis
T. rangeli/T. conorhini
AF301912 Trypanosoma cruzi
AJ009151 Trypanosoma dionisii
Schyzotrypanum
KT030823 Trypanosoma wauwau
T. wauwau
T. cruzi clade
79/73
100/100
65/92
100/100
70/78
100/100
53/100
100/100
100/100
77/71
88/-
80/94
100/100
99/100
100/100
100/100
54/96
84/100
77/-
80/97
81/79
96/98
95/92
100/100
99/100
100/100
61/80
83/69
100/100
100/100
55/52
100/100
99/100
100/100
69/53
64/-
79/-
84/98
78/94
100/100
0.1
ML/NJ

# Secondary Structure: small subunit ribosomal RNA




*Trypanosoma brucei*
(M12676)
1.cellular organisms 2.Eukaryota 3.Euglenozoa
4.Kinetoplastida 5.Trypanosomatidae
6.Trypanosoma 7.Trypanozoon
July 2003