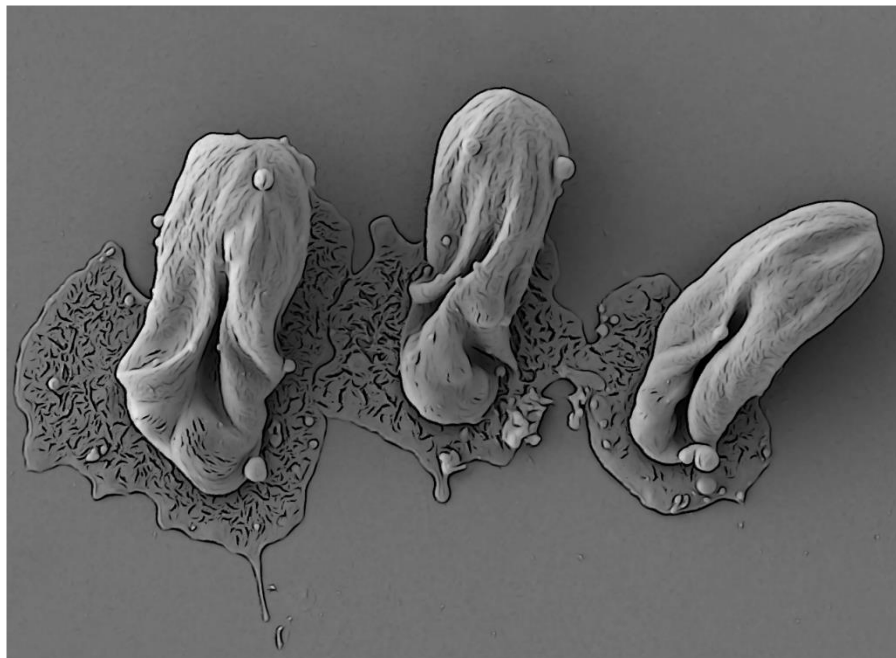
School of Doctoral Studies in Biological Sciences

UNIVERSITY OF SOUTH BOHEMIA, FACULTY OF SCIENCE

Insight into insect trypanosomatid biology via whole genome sequencing

Ph.D. Thesis

RNDr. Tomáš Skalický



Supervisor: Prof. RNDr. Julius Lukeš, CSc.

Biology Centre CAS, Institute of Parasitology

České Budějovice 2017

This thesis should be cited as:

Skalický T (2017) Insight to insect trypanosomatid diversity via whole genome sequencing. Ph. D. Thesis. Uniccersity of South Bohemia, Faculty of Science, School of Doctoral Studies in Biological Science, České Budějovice, Czech Republic, 94 pp.

Annotation

This thesis is composed of two topics both concerning diverse and obligatory trypanosomatid parasites. First part deals with identification of new *Trypanosoma* species identified in blood meal of tsetse flies caught in Dzanga-Sangha Protected Areas, Central African Republic, and identification of feeding preferences of tsetse flies. The second part concerns extraordinary monoxenous trypanosomatid *Paratrypanosoma confusum* which constitutes the most basal branch between free-living *Bodo saltans* and parasitic trypanosomatids. This thesis helped to elucidate morphology and biology of this deep branching trypanosomatid. Using genome and transcriptome sequencing and comparative bioinformatics approaches enabled search for ancestral genes shared with free-living bodonids and confirmed genome streamlining in trypanosomatids.

Declaration [in Czech]

Prohlašuji, že svoji disertační práci jsem vypracoval samostatně pouze s použitím pramenů a literatury uvedených v seznamu citované literatury.

Prohlašuji, že v souladu s § 47b zákona č. 111/1998 Sb. v platném znění souhlasím se zveřejněním své disertační práce, a to v nezkrácené podobě elektronickou cestou ve veřejně přístupné části databáze STAG provozované Jihočeskou univerzitou v Českých Budějovicích na jejích internetových stránkách, a to se zachováním mého autorského práva k odevzdanému textu této kvalifikační práce. Souhlasím dále s tím, aby toutéž elektronickou cestou byly v souladu s uvedeným ustanovením zákona č. 111/1998 Sb. zveřejněny posudky školitele a oponentů práce i záznam o průběhu a výsledku obhajoby kvalifikační práce. Rovněž souhlasím s porovnáním textu mé kvalifikační práce s databází kvalifikačních prací Theses.cz provozovanou Národním registrem vysokoškolských kvalifikačních prací a systémem na odhalování plagiátů.

České Budějovice, 06.07.2017 ....

Tomáš Skalický

This thesis originated from a partnership of Faculty of Science, University of South Bohemia, and Institute of Parasitology, Biology Centre of the CAS, supporting doctoral studies in the Molecular and Cell Biology and Genetics study programme.

Přírodovědecká
fakulta
Faculty
of Science

BIOLOGICKÉ
CENTRUM
AV ČR, v. v. i.

Financial support

This work was supported by:

ERC CZ, grant LL1205

The Czech Science Foundation, grant 14-23986S,

The Praemium Academiae 2009 award to Julius Lukeš

The Grant Agency of the University of South Bohemia, grant 108/2013/P

Acknowledgements

Foremost, I would like to express my sincere gratitude to my supervisor Julius Lukeš for giving me the opportunity to join his team, his patient guidance, motivation and continuous support of my research during all the years. His enthusiasm and passion for science was always contagious and brought me where I stand now. I would also like to thank all the people from the institutes that I had the opportunity to work with during the years, especially those from Jula's, Alča's and Parda's lab, both former and present. They have created a pleasant and stimulating working environment throughout my PhD studies. I would also like to thank David Wildridge and Carmen Priscila Pena Diaz for proofreading.

In addition, my thanks go to all my friends, whom I had the chance to get to know over the years spent in South Bohemia. Thanks to all of them, my studies on Faculty of Sciences were both happy and pleasurable.

Finally, my special thanks go to my beloved Zuzka and my family for their never-ending support, help and constant faith in me.

List of papers and author's contribution

The thesis is based on the following papers (listed chronologically):

I. Flegontov P*, Votýpka J*, **Skalický T***, Logacheva MD, Penin AA, Tanifuji G, Onodera NT, Kondrashov AS, Volf P, Archibald JM, Lukeš J (2013) *Paratrypanosoma* is a novel early-branching trypanosomatid. **Current Biology** 23(18):1787-93 (IF = 9.733).

Tomáš Skalický participated in study design, experimental data generation, data analysis and interpretation and draft editing of the manuscript.

II. Lukeš J, **Skalický T**, Týč J, Votýpka J, Yurchenko V (2014) Evolution of parasitism in kinetoplastid flagellates. **Molecular and Biochemical Parasitology** 195(2):115-122 (IF = 2.068).

Tomáš Skalický performed analysis and interpretation of the data, prepared figures and contributed to writing of the manuscript.

III. Votýpka J*, Rádrová J*, **Skalický T***, Jirků M, Jirsová D, Mihalca AD, D'Amico G, Petrželková KJ, Modrý D, Lukeš J (2015) A tsetse and tabanid fly survey of African great apes habitats reveals the presence of a novel trypanosome lineage but the absence of *Trypanosoma brucei*. **International Journal of Parasitology** 45(12):741-8 (IF = 4.242).

Tomáš Skalický participated in experimental procedures, designed, prepared and sequenced amplicon DNA library, analyzed the results, performed recombinant detection and contributed to writing of the manuscript.

IV. **Skalický T***, Dobáková E*, Wheeler R*, Tesařová M, Flegontov P, Jirsová D, Votýpka J, Yurchenko V, Ayala FJ, Lukeš J (submitted manuscript) Extensive flagellar remodeling during the complex life cycle of *Paratrypanosoma*, an early-branching trypanosomatid.

Tomáš Skalický participated in study design, experimental data generation, genome assembly and annotation, data analysis and interpretation and contributed to writing of the manuscript.

* These authors contributed equally

Co-author agreement

Julius Lukeš, the supervisor of this Ph.D. thesis and co-author of all presented papers, fully acknowledges the contribution of Tomáš Skalický.

…………………………………………

Prof. RNDr. Julius Lukeš, CSc.

Contents

1 Introduction ... 1
1.1 Kinetoplastea ... 1
1.2 Trypanosomatida ... 2
1.2.1 Trypanosomatid systematics and phylogeny ... 3
1.2.2 Trypanosomatid diversity ... 5
1.2.3 Evolution of parasitism in trypanosomatids ... 10
1.2.4 Trypanosomatid nuclear genomes ... 13
1.2.5 RNA interference ... 16
2 Research objectives ... 19
3 Summary of results and discussion ... 19
3.1 Survey of new trypanosomes transmitted by tsetse fly ... 20
3.2 *Paratrypanosoma confusum,* a unique trypanosomatid ... 21
4 REFERENCES ... 27
5 Original publications ... 36
5.1 PAPER I ... 36
5.2 PAPER II ... 55
5.3 PAPER III ... 64
5.4 MANUSCRIPT I ... 73
6 Curriculum vitae ... 89

1 Introduction

1.1 Kinetoplastea

Kinetoplastids represent one of the putatively most ancestral eukaryotic lineages (Cavalier-Smith, 2010). This widespread and extremely speciose group of eukaryotic protists, sometimes called microeukaryotes, belongs to the Euglenozoa, supergroup Excavata (Adl et al., 2012). Together with three sister groups Euglenida, Symbiontida and Diplonemea, Kinetoplastid flagellates are characterized by specific features including polycistronic transcription and *trans*-splicing of nuclear genes or a large mitochondrion with discoidal cristae containing a prominent mitochondrial DNA termed kinetoplast, which gave the group its name.

Phylogenetic analyses support division of the class Kinetoplastea into subclasses Prokinetoplastina and Metakinetoplastina (Adl et al., 2012; d'Avila-Levy et al., 2015), with the former clade containing only two known representatives: *Ichthyobodo* – an ectoparasite of fish with two flagella and *Perkinsela* – an aflagellar endosymbiont of amoebozoans of the genus *Paramoeba* parasitizing on gills of fish (Dykova et al., 2008; Young et al., 2014). *Perkinsela,* which according to the Giemsa staining has more DNA in its mitochondrion than in its nucleus, is located in the perinuclear region of the host and seems to behave like an organelle, as it synchronizes its division with the division of the host (David et al., 2015; Tanifuji et al., 2011).

All the remaining species belong to subclass Metakinetoplastina, which is further subdivided into four clades – Eubodonida, Neobodonida, Parabodonida, and Trypanosomatida. The first three groups are called bodonids and comprise species with various life styles ranging from free-living, through commensalic to parasitic ones. Phylogeny of this group is still not fully resolved, with free-living genera *Bodo* and *Neobodo* being paraphyletic (Lukeš et al., 2014). Parasitic life style has apparently evolved several times within bodonids, and currently there are three known parasitic clades; snail-infecting *Cryptobia* (Woo, 2003), economically important *Azumiobodo*, which causes severe damage on cultured ascidians (Kumagai et al., 2013), and *Trypanoplasma* living in the fish blood (Lukeš et al., 1994). The last group – Trypanosomatida contains remarkably successful obligatory parasites infecting a wide range of invertebrates, vertebrates and plants (see below).

1.2 Trypanosomatida

Trypanosomatids are probably the most studied group of protists (Verner et al., 2015), with species that are present both globally and endemically (Votýpka et al., 2012b). In addition to features common for most kinetoplastids (one large mitochondrion, RNA editing, polycistronic transcription and *trans*-splicing of nuclear genes, compartmentalized glycolysis, etc.), trypanosomatids possess additional features that are characteristic solely for this group. These include: i/ fully developed subpellicular microtubular corset, which reinforces the cell and at the same time restricts endo- and exocytosis to the flagellar pocket (Hemphill et al., 1991; Vickerman, 2000), ii/ single beating flagellum that can be either free or attached to the cell body, creating an undulating membrane (Vickerman, 2000), iii/ paraflagellar rod serving as a support for external part of the flagellum (Vickerman, 1962), iv/ acidocalcisomes with high concentration of calcium for ion homeostasis regulation, and v/ a compactly packed kinetoplast (k) DNA, which is a complex of interconnected circular DNA molecules of two categories: maxicircles and minicircles (Liu et al., 2005).

In contrast to their paraphyletic sister group Bodonida, all members of the family Trypanosomatidae are obligatory parasites and many of them are of high medical and veterinary importance (Lukeš et al., 2014; Simpson et al., 2006). Based on their life cycle, we can distinguish one-host (monoxenous) parasites infecting various insect species, living predominantly in their digestive tract and two-host (dixenous) parasites that added a second host – either a plant or more commonly a vertebrate or invertebrate host (Kaufer et al., 2017; Maslov et al., 2013); see below for details). Trypanosomatids are morphologically very variable; currently eight morphotypes are recognized (epimastigote, trypomastigote, promastigote, opisthomastigote, choanomastigote, amastigote, spheromastigote and opisthomorph) (Fig. 1). These are defined by cell morphology, mutual positioning of the kinetoplast, nucleus and flagellar pocket and presence/absence of the flagellum (d'Avila-Levy et al., 2015; Wallace et al., 1983).

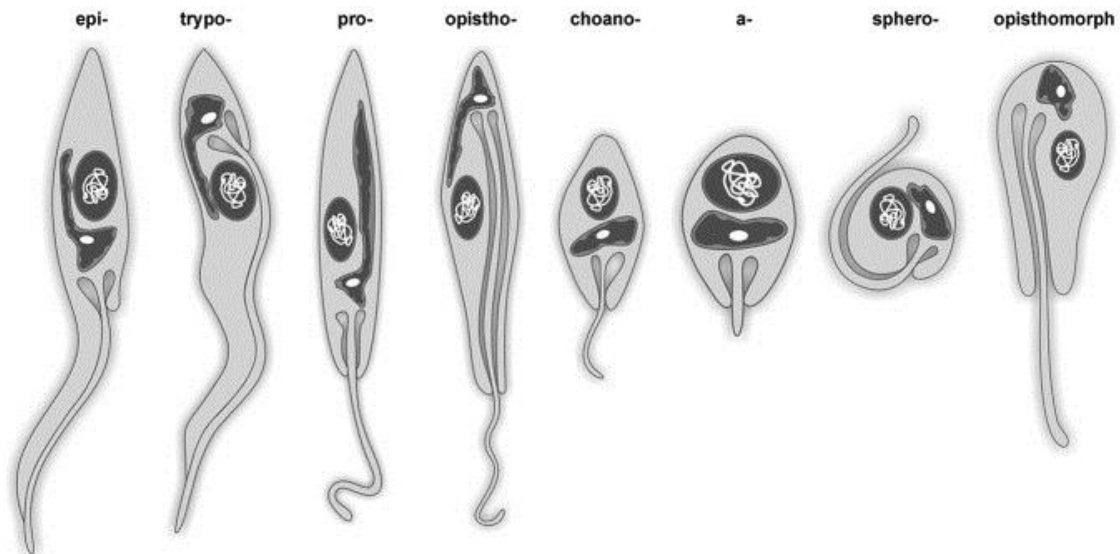


Figure 1: Schematic representation of the main morphological forms present in trypanosomatids. Dash should be replaced by the word "mastigote" (Adapted from d'Avila-Levy et al., 2015).

1.2.1 Trypanosomatid systematics and phylogeny

Until the end of 19th century, trypanosomatid taxonomy was based on the presence of a set of life cycle stages, disease manifestation and morphology using light microscopy (Hoare and Wallace, 1966; Simpson et al., 2006). Later usage of electron microscopy techniques did not yield any significant advancement for species recognition and identification, but allowed detailed description of the flagellar pocket, cytoskeletal organization and kinetoplast structure (De Souza, 1995; Gluenz et al., 2015; Wheeler et al., 2016). Until the end of 20th century, only about a dozen trypanosomatid genera has been described, mainly from medically important clades (Simpson et al., 2006). This is not surprising considering that trypanosomatids have a predominantly clonal reproduction and lack sufficient morphological differences among species. Moreover, even within a single isolate, there is a continuum of morphotypes overlapping with those of other isolates (Maslov et al., 2013). Thanks to major advancements in molecular biology, DNA sequencing and phylogenetic approaches, our view of trypanosomatid phylogeny has improved, however due to an uneven sampling, with focus on medically important species, our view remains deeply biased.

Currently, several molecular markers have been extensively used. Probably the most widely used and information-rich one with the highest number of representative sequences in online databases is the nuclear small subunit (SSU, 18S) rRNA. Unfortunately, relying solely on this marker has numerous limitations. This gene has gone through a massive

evolutionary change in the early kinetoplastids and, therefore, phylogenetic trees have extremely long branches that connect kinetoplastids with other eukaryotes, yet at the same time create short internal branches within this group (Maslov et al., 1996; Simpson et al., 2006). To overcome this limit, other genetic loci must be used. The second most frequently used marker is the glycosomal glyceraldehyde 3-phosphate dehydrogenase gene that appears to be a good choice for building phylogenetic trees, particularly for these protists (Hollar et al., 1998). In a limited number of cases the usage of other housekeeping genes, such as nuclear large subunit (LSU, 28S) rRNA, RNA polymerase II or heat-shock proteins hsp70 and hsp83 also proved to be helpful (Barratt et al., 2017; Kostygov et al., 2016; Pothirat et al., 2014). Nowadays, it is obvious that a single gene-based phylogenetic analyses often provide inaccurate tree topologies with poorly supported branches, because sequencing is prone to errors and there are not enough informational positions in one gene. Therefore, using concatenated phylogenetic matrices combining at least one rRNA gene, other nuclear and mitochondrial-encoded genes are becoming standard tools required for comprehensive analyses. Combination of these markers usually provides well-supported inter-specific resolution but because of their relative sequence conservation, they fail to resolve terminal branches of the phylogenetic trees that represent closely related species or strains of the same species.

For this purpose, one can make advantage of the internal transcribed spacers (ITSs) 1 and 2 and kinetoplastid-specific splice leader (SL) RNA gene repeats (Maslov et al., 2013; Votýpka et al., 2014). ITS1 gene is located between 18S and 5.8S rRNA genes, while ITS2 can be found between 5.8S and 28S rRNA genes. These genes are easy to amplify because of their relatively small size (cca 500 nucleotides [nt]) and highly conserved flanking sequences. They are also useful for detection from very small amounts of DNA because they are present in high copy number in the genomes.

The SL RNA marker combines three types of sequences: a highly conserved 39 nt mini-exon, 50-60 nt moderately conserved intron and a highly variable intergenic sequence. The latter is responsible for length variability of the SL repeats, which range from 0.2 kb to 1.0 kb in different trypanosomatids and are tandemly arranged as blocks of 100-200 copies in their genomes. This combination of conserved and variable regions enabled development of a PCR-based approach, wherein the conserved exon sequences served as a target for universal primers designed to amplify an entire repeat unit, while the hypervariable intergenic region was used as a molecular marker of high resolution (Campbell et al., 2000; Teixeira et al., 2011).

When combined with the traditional morphology-based system, these modern phylogenetic techniques brought major changes into the classification of trypanosomatids, in numerous cases leading to the redefinition of genera. New nucleotide sequence-based system substituting species with typing units (TUs) was successfully tested and then introduced (Kozminsky et al., 2015; Maslov et al., 2013; Westenberger et al., 2004). This approach defines a "molecular species" with 90% sequence similarity threshold of the SL RNA gene repeats, identifying individual TUs.

Combination of phylogenetic analyses based on concatenated datasets which are supported by the SL analysis was tested on extensive field samples from Costa Rica and Ecuador (Yurchenko et al., 2008), and is now routinely used for field surveys and species descriptions (Barrat et al., 2017; Espinosa et a., 2016; Kozminsky et al., 2015; Schwarz et al., 2015; Svobodová et al., 2007; Votýpka et al., 2010). For example, studies of trypanosomatid parasites of Neotropical Heteroptera conducted in South-West China, central Europe, the Mediterranean and equatorial Africa revealed more than 90 TUs representing new species according to the SL-based genotyping (Jirků et al., 2012; Kostygov et al., 2014; Votýpka et al., 2012a; Votýpka et al., 2010).

These modern techniques also proved that trypanosomatids really evolved from bodonids (Simpson et al., 2006), disproved the "one host – one parasite" paradigm (Podlipaev, 1990; Votýpka et al., 2015), supported monophyly of the genus *Trypanosoma* (Hamilton et al., 2004; Leonard et al., 2011) and brought attention to the monoxenous species that were until present, overlooked because of their "insignificance".

1.2.2 Trypanosomatid diversity

Until the 1990s, the collection of monoxenous species was very small, with only a few new species being added to the organisms kept in culture already for decades. However, the recent emergence of several new major clades or independent deep-branching lineages within the monoxenous trypanosomatids triggered a new wave of interest and motivated extensive sampling for these insect parasites around the world. Consequently, almost 400 insect trypanosomatids have been described to date, and this number is steadily increasing (d'Avila-Levy et al., 2015; Kostygov et al., 2016; Kozminsky et al., 2015; Kraeva et al., 2015; Podlipaev et al., 2004; Týč et al., 2013; Votýpka et al., 2014).

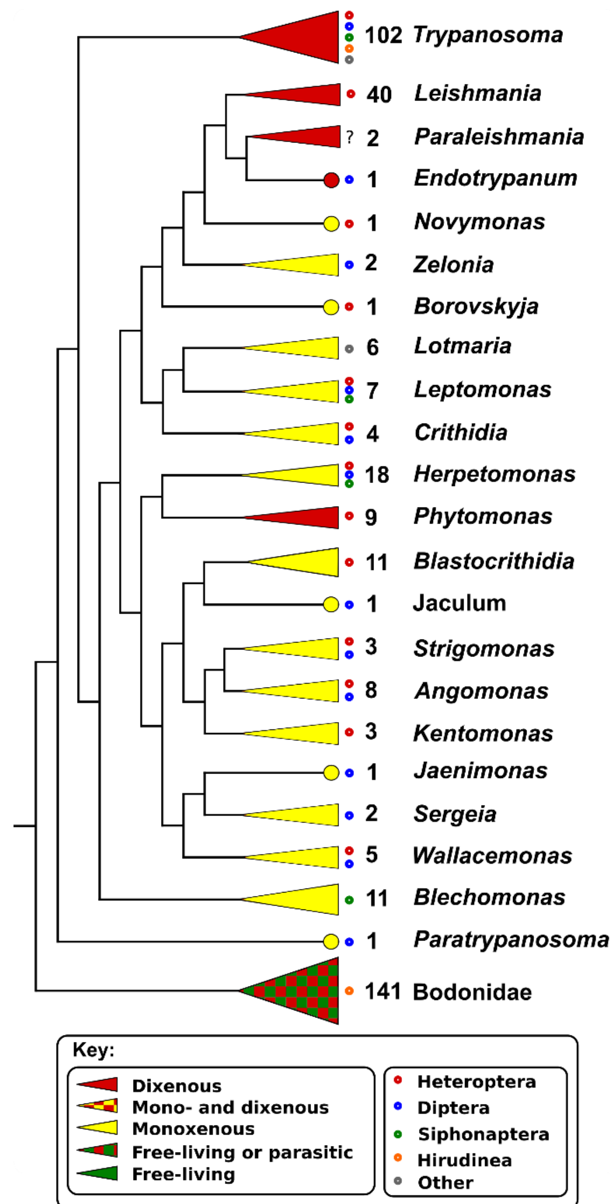


Figure 2: Evolutionary tree of trypanosomatids based on molecular data. Number of sequences representing species are shown. Red-green checkboard represents free-living and parasitis bodonids species collapsed in one clade. Yellow color represents monoxenous species and red color dixenous. Colored circles depict groups of invertebrate hosts. (Adapted with updates from Lukeš et al., 2014)

According to the latest trypanosomatid systematics, 22 major clades have been identified (Fig. 2), which is double of those known just ten years ago (Kaufer et al., 2017; Simpson et al., 2006). Although the relationships among them are not resolved with high support and some genera are still being disputed (Kostygov and Yurchenko, in press), most of the subdivisions are robust and well-supported. They divide the described species or TUs into meaningful groups, which are reflected in their life styles, hosts, or molecular features (Kaufer et al., 2017; Lukeš et al., 2014; Maslov et al., 2013). Still, because of their medical and economical importance, the best known and most studied members belong to the

dixenous genera *Trypanosoma, Leishmania* and *Phytomonas.* There are more than 140 described species with sequence data available in NCBI database.

The genus *Trypanosoma* is the most species-rich. Its members parasitize a wide range of hosts encompassing mammals, fish, birds, amphibians and lizards (Podlipaev, 1990), and are transmitted by vectors ranging from leeches (Hayes et al., 2014) to biting flies and bugs (Franco et al., 2014; Stevens et al., 1999). In the case of tsetse flies, it was thought that these transmit only *Trypanosoma brucei,* but modern high-throughput sequencing methods revealed that they actually act as vectors for a wide range of trypanosomes (Votýpka et al., 2015). PCR-based analysis of their blood prey even allowed determining on which vertebrate species a given tsetse fly was feeding (Calvignac-Spencer et al., 2013). Indeed, these data may be also used to sample local vertebrate biodiversity, including very rarely seen vertebrates (Calvignac-Spencer et al., 2013).

The most pathogenic and at the same time best studied *T. brucei* is further subdivided into five sub-species, each inflicting different form of trypanosomiasis. *T. brucei brucei* causes African trypanosomiasis, also called nagana (Roberts et al., 2013) is non-infectious to humans because of its susceptibility to lysis by human apolipoprotein L1 (Pays and Vanhollebeke, 2008), and hence infects wild African animals (antelopes, zebras and others) and domestic animals (horses, goats, pigs and others).

T. brucei equiperdum and *T. brucei evansi* have partially (dyskinetoplastidy) or totally (akinetoplastidy) lost their kinetoplast DNA (Lai et al., 2008). *T. brucei equiperdum* infects horses and causes a deadly disease known as dourine. These trypanosomes can mainly be found in the lymph and skin blisters, and are transmitted mechanically during copulation. *T. brucei equiperdum* can be found outwith Africa, including areas of Asia and Eastern Europe (Volf and Horák, 2007). *T. brucei evansi* infects camels and water buffaloes, causes a serious disease called surra, which is transmitted both by bloodsucking insects and during mating, and occurs worldwide (Lai et al., 2008).

The most medically important trypanosomes are *T. brucei gambiense* and *T. brucei rhodesiense.* The former can be found in western and central parts of Africa and inflicts so-called West African or chronic sleeping sickness in humans. This disease may take several years to manifest but is eventually deadly if left untreated. *T. brucei rhodesiense* is most common in southern and eastern savannah regions, infecting wild animals, livestock and humans. In humans, it causes so-called East African or acute sleeping sickness, which is deadly in several weeks if left untreated (Barrett et al., 2003; Volf and Horák, 2007). Human

African Trypanosomiasis caused by these parasites is endangering up to 61 million people, with approximately 15,000 new cases every year (Kaufer et al., 2017).

These obligate parasites, except *T. brucei equiperdum* and *T. brucei evansi,* that exist only in the bloodstream form and are transmitted mechanically (Brun et al., 1998), have two hosts – an insect vector (tsetse fly *Glossina* species) and a mammalian host. Differences between these hosts are substantial in every aspect, and to cope with switching between them, during its life cycle trypanosome undergoes complex changes in morphology, metabolism and surface proteins. The main stages are the bloodstream stage in mammalian blood, the procyclic stage in the tsetse fly midgut, and the infectious metacyclic stage in the tsetse fly salivary glands (Verner et al., 2015; Vickerman, 1985).

Another member of this genus is *Trypanosoma cruzi* that causes Chagas disease, an illness endemic to Latin America with an estimated eight million people infected. This parasite is transmitted by a blood-sucking reduviid bug of the Triatominae subfamily. *T. cruzi* is primarily a zoonotic parasite with dogs, cats and wildlife serving as its reservoirs (Jackson et al., 2014). Humans are incidental hosts. The infectious stages of *T. cruzi*, called metacyclic trypomastigotes, are located in the rectal cell wall of the bug and are transmitted by defecation on skin of vertebrate host during blood feeding (Rassi et al., 2010).

Protists ranked into the genus *Leishmania* seriously affect millions of people in tropical and subtropical regions of the world. These parasites are transmitted by females of sandflies of the genera *Phlebotomus* and *Lutzomyia* during the blood meal, and are causative agents of different kinds of leishmaniases (Alvar et al., 2012; Banuls et al., 2007). These intracellular parasites cause three different forms of infections: cutaneous leishmaniasis (caused by *L. major, L. tropica, L. aethiopica* and *L. mexicana*), mucocutaneous leishmaniasis (caused by *L. brasiliensis*) and visceral form (caused by *L. donovani* and *L. infantum*) (Ryan and Ray, 2003).

Representatives of the genus *Phytomonas* are parasites of a wide range of plants and can be divided into three groups according to their occurrence in floem, latex ducts, fruit and seeds. They are transmitted by phytophagous insects and are responsible for deadly diseases of economically important crops, such as coffee, corn, coconut, cassava, tomatoes and hartrot of oil palms (Camargo, 1999; Jaskowska et al., 2015).

Another dixenous genus is *Paraleishmania* that was recently re-erected and contains two species formerly known as *Leishmania hertigi* and *Porcisia deanei* (Kostygov and Yurchenko, in press). Both species were isolated from rodents in South America and their

vectors are currently unknown (Espinosa et al., 2016). *Endotrypanum* is the last known dixenous trypanosomatid infecting erythrocytes of two-toed sloth (Mesnil and Brimont, 1908) with its vector never identified, but phlebotomine sand flies are suspected (Shaw, 1964). This genus is highly controversial and its identity has been debated for decades (Cupolillo et al., 2000). All currently isolated strains that were presumed to belong to it lack the originally described morphology and their ability to infect erythrocytes was never proven. Therefore, it was proposed to consider this genus as *nomen dubium* until these issues have been resolved (Kostygov and Yurchenko, in press).

Regarding the monoxenous species and their diversity, they are limited only by the number of possible niches and the evolutionary time needed for diversification. Even if one would assume that not every insect species is suitable for new niche establishment, as the most diverse group of eukaryotes, insects represent an extremely speciose group of hosts. Scientists have been able to identify almost one million of insect species and the total number is estimated over five million (Stork et al., 2015). From those known to us, only 2500 have been examined for the presence of trypanosomatids, and even this small fraction significantly extended our knowledge of trypanosomatids and unveiled the facts that most of trypanosomatid diversity lays within the monoxenous species.

At the time of writing this thesis, 17 monoxenous clades have been recognized. *Angomonas, Kentomonas, Strigomonas* and *Novymonas* are bacterial endosymbiont-bearing clades with *Novymonas* hosting a bacterium not related to other trypanosomatid endosymbionts. Moreover, it seems that the *Novymonas* symbiosis is a relatively recent event as the host does not fully control the number of harbored bacteria (Kostygov et al., 2016). The genera *Crithidia, Herpetomonas* and *Leptomonas* are polyphyletic according to recent phylogenetic analyses (Votýpka et al., 2015; Yurchenko et al., 2009). Indeed, substantial differences among species belonging to the same clade or genus were already noticed in the past, e.g. during comparison of enzymatic activities and DNA hybridization surveys (Batistoti et al., 2001). As a matter of fact, *Leptomonas* and *Crithidia* have been considered as the lumping genera for all species with unresolved phylogenetic position (Batistoti et al., 2001; Yurchenko et al., 2008). Consequently, it is not surprising that several species from the genus *Crithidia* were recently reassigned into the newly erected genus *Lotmaria,* which brings together species parasitizing honey bees (Schwarz et al., 2015). Several new genera have been established for species separated from the genus *Leptomonas.* One of them is *Leptomonas barvae*, which together with numerous environmental samples justified constitution of the genus *Borovskyja* (Kostygov and Yurchenko, in press). Another

genus was erected by reclassification of *Leptomonas costaricensis,* constituting the genus *Zelonia* which also encompasses several newly identified species from Australia (Barratt et al., 2017).

An interesting case of a taxonomy mix-up is represented by the genus *Wallacemonas,* which was previously considered polyphyletic. The reason was that the type species was actually a mixed culture of two morphologically very different species, one being true *Wallacemonas* and the second one belonging to the genus *Crithidia,* which led to contradictory results between taxonomy and molecular phylogeny analyses (Kostygov et al., 2011; Kostygov et al., 2014). Species from the clade "Jaculum" and the genus *Blastocrithidia* are difficult to cultivate, requiring complex cultivation media. The latter one got its share of publicity when it was found it uses a non-canonical nuclear genetic code for all three termination codons, where UGA has been reassigned to encode tryptophan, while UAG and UAA (UAR) encode glutamate, with only UAA serving as a genuine stop (Záhonová et al., 2016). It seems that some trypanosomatid species infect only certain group of insects. For example, representatives of the genus *Sergeia* are confined to biting midges (Svobodová et al., 2007); while fruit flies (*Drosophila* sp.) are being infected by species from the *Jaenimonas* clade (Hamilton et al., 2015), and fleas harbor members of the genus *Blechomonas* (Votýpka et al., 2013). Last but not least, is the genus *Paratrypanosoma*, which forms the most basal branch between free-living bodonids and parasitic trypanosomatids (Flegontov et al., 2013) (described in more detail in chapter 3.2).

1.2.3 Evolution of parasitism in trypanosomatids

The scientific community has been struggling to unravel the evolution of trypanosomatids, sometimes producing contradictory reports concerning their phylogenetic relationships. Two mutually exclusive scenarios of the origin of dixenous (two-host) trypanosomatids are being debated. The 'vertebrate-first' suggests that the ancestral flagellates first colonized the gut, only later the blood of early aquatic vertebrates, and were subsequently introduced to the haematophagous insects during blood feeding (Minchin, 1908). The 'insect-first' hypothesis postulated that parasites of vertebrates descended from those of insects, where they became established first (Léger, 1904). The struggle between these evolutionary scenarios lasted for almost a century because until recently, the most basal trypanosomatid branch was constituted by the genus *Trypanosoma,* supporting Minchin's vertebrate-first hypothesis. Thanks to intensive sampling covering a greater area and

improved modern molecular phylogenomic methods, a new basal branching monoxenous trypanosomatid named *Paratrypanosoma confusum* has been described. The available data robustly verify its position at the transition between free-living bodonids and dixenous *Trypanosoma,* thus ultimately proving Léger's hypothesis that all dixenous flagellates are derived from their monoxenous predecessors from the insect hosts (Flegontov et al., 2013; Lukeš et al., 2014; Maslov et al., 2010). This course of events had additional support from an unexpected fossil source. A phlebotomine fly trapped in amber and massively infected by a flagellate highly reminiscent by shape of *Leishmania* was identified in Myanmar. The parasite labeled *Paleoleishmania*, was accompanied in the insect's intestinal tract by nucleated red blood cells, likely originating from a dinosaur (Poinar, 2008). *Paleoleishmania,* morphologically virtually indistinguishable from the extant *Leishmania,* was dated to ~110 million years, indicating that the establishment of the dixenous life style might be a fairly ancient event (Poinar, 2008; Poinar and Poinar, 2004). Modern molecular dating analyses using multi-concatenated protein datasets built from dozens of available trypanosomatid nuclear genomes place the origin of the genera *Leishmania* and *Trypanosoma* at least 90-100 MYA (million year ago) (Fig. 3) (Harkins et al., 2016; Lukeš et al., 2014). This is quite plausible at least for the *Trypanosoma* species, where the divergence between *T. brucei* and *T. cruzi* is believed to have occurred 100 MYA based on the split between Africa and South America (Lukeš et al., 2007; Stevens and Gibson, 1999). Nonetheless, the phylogeographic data has to be treated with caution as was shown in the case of alligators and crocodiles. These species split around 90 MYA, although trypanosomes parasitizing them are still closely related, a finding more consistent with a transoceanic dispersal of *Crocodylus* around 4-5 MYA (Fermino et al., 2013). Moreover, we need exercise caution when using molecular dating, as calibration points in the form of fossils are missing; the further we go into the past, the higher the uncertainty.

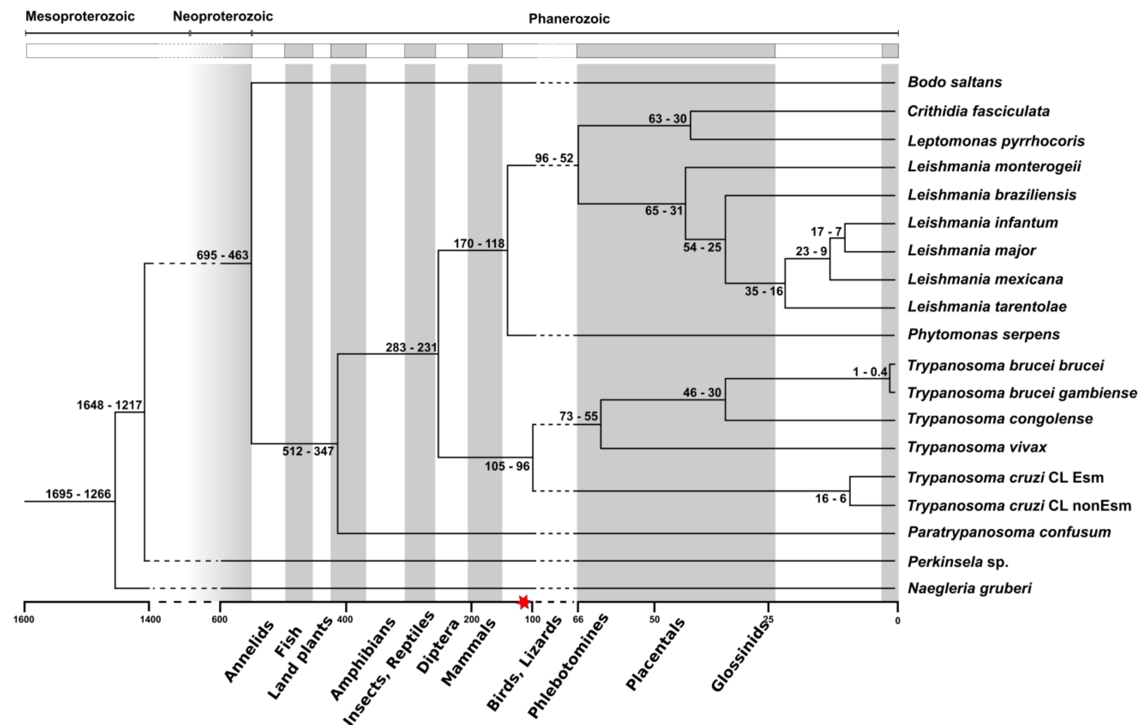


Figure 3: Phylogenetic tree of kinetoplastids constructed using molecular clock model. Concatenated dataset of 42 proteins for 18 kinetoplastid species and *Naegleria* serving as an outgroup was used to estimate divergence times. Node represents mean divergence and numbers 95% confidence intervals. Star denotes oldest known fossil of a kinetoplastid. X axis is in absolute time scale in millions of years along with emergence of host groups invaded by trypanosomatids. (Adapted from Lukeš et al., 2014).

There are reports that monoxenous trypanosomatids occasionally explore the potential of a novel niche. While being considered non-pathogenic for vertebrates, monoxenous trypanosomatids were occasionally found infecting immunodeficient humans. The capacity to invade vertebrate host is not restricted to one group, as there are several unrelated isolates that possess this ability. For example, a close relative of *Blechomonas pulexsimulantis* from the monoxenous genus normally infecting fleas was isolated from a HIV-positive person in Brazil (Pacheco et al., 1998). Several other cases of HIV-positive individuals co-infected with a monoxenous trypanosomatid from various genera have been reported over the years (Chicharro and Alvar, 2003). It seems that these patients may develop skin lesions, splenomegaly and other symptoms typical for visceral or cutaneous leishmaniasis (Barratt et al., 2010).

Not surprisingly, monoxenous trypanosomatids from the genus *Leptomonas* also have been detected as co-infections in HIV-negative patients suffering from visceral leishmaniasis or kala-azar caused by *Leishmania donovani* (Selvapandiyan et al., 2015; Singh et al., 2013; Srivastava et al., 2010). The main monoxenous species recovered from these co-infections

is *Leptomonas seymouri* (Singh et al., 2013) originally isolated from a "cotton stainer" *Dysdercus suturellus,* a strictly phytophagous bug (Wallace, 1977). Humans are not the only vertebrates which are being probed for a novel niche by monoxenous trypanosomatids, as members of the genus *Herpetomonas* were also isolated from rats and stray dogs in Egypt (Morsy et al., 1988; Podlipaev et al., 2004).

1.2.4 Trypanosomatid nuclear genomes

Trypanosomatids undoubtedly belong to the most successful parasites on Earth with unique and dexterous life style and almost ingenious adaptations to the hostile environment of the host. Because they cause devastating diseases, several *Trypanosoma* and *Leishmania* species have been subjected to a great number of studies aiming to understand various aspects of their molecular and cellular biology. Many fascinating features have ben found over the last decades. One such finding is that trypanosomatids have virtually intron-less genomes. The only exception is polyA polymerase and putative DEAD/H RNA-helicase, which possess a *cis*-spliced intron (Mair et al., 2000; Preusser et al., 2014).

Trypanosomatids also perform polycistronic transcription, an extremely rare feature in eukaryotes. Trypanosomatid genes are transcribed in the form of very long transcripts, sometimes encompassing the entire length of a chromosome (Martinez-Calvillo et al., 2003; Worthey et al., 2003). Transcription is initiated by polymerase II, usually in bidirectional fashion and the emerged transcripts are subsequently processed into monocistronic units. Capped 5'-end of mRNA is generated by *trans*-splicing with the addition of a 39 nt-long capped SL RNA, while polyadenylation of the 3'-end is coupled with *trans*-splicing of the downstream unit (Clayton, 2016; LeBowitz et al., 1993). All necessary steps are performed in the nucleus by the splicesome (Mayer and Floeter-Winter, 2005). Polycistronic transcription means that trypanosomatids are unable to regulate expression of individual genes, and the regulation is post-transcriptional. There are several possible methods to regulate transcription in the nucleus prior to mature mRNAs reaching the cytosol. They range from varying the nuclear gene copy number (Clayton, 2016), through competition between mRNA processing and RNA degradation machineries (Fadda et al., 2014), to transcription unit length and gene order (Kelly et al., 2012).

The rapid development of high-throughput sequencing techniques allowed genome and transcriptome sequencing of dozens of medically and economically important dixenous parasites, producing high quality assemblies that can serve as a suitable reference. Genome

sizes of sequenced trypanosomes range from 22 Mb in *T. brucei gambiense,* through 26 Mb in *T. brucei brucei* (strain TREU927), to 47.5 Mb in *T. vivax* (Gibson, 2012; Greif et al., 2013; Sistrom et al., 2014). The genomes of *Leishmania* species are more comparable in size, ranging from 27 Mb in *Leishmania enriettii* to 33 Mb in *L. infantum* and *L. major* (Peacock et al., 2007; Real et al., 2013). All other genomes for dixenous parasites belong to the genus *Phytomonas,* namely *Phytomonas sp.* strains EM1 and HART with genome size of 18.1 Mb and 17.8 Mb, respectively (Porcel et al., 2014). Quite recently a 17.7 Mb genome of *Phytomonas françai* was announced (Butler et al., 2017).

Initial comparative genome analysis of *T. brucei, T. cruzi* and *L. major,* frequently labeled as "Tritryps", lead to several unexpected findings (El-Sayed et al., 2005). Haploid genomes of Tritryps are divided into 11 to 36 chromosomes, which contain from 8100 to 12000 protein-coding genes (El-Sayed et al., 2005). Although these pathogens substantially differ by their insect vectors, target tissues, immune evasion strategy and disease pathogenesis, they share a conserved proteome core of almost 6200 genes. Moreover, 94% of them belong to regions of conserved gene syntheny (El-Sayed et al., 2005). Currently, about 50% of trypanosomatid genes have no known function (Jackson, 2015). Most of them do not have homologues outside of Kinetoplastea, so a tedious functional analysis is the only way to establish function, which has a potential in the development of new chemotherapeutics, with the advantage of their likely use against all three parasites.

The fact that the Tritryp species are so dissimilar in their pathology and at the same time share so many features on the genome level is even more striking when comparing different species within a genus. For example, genome-wide comparative analyses of three *Leishmania* species, namely *L. major, L. infantum* and *L. braziliensis* that are clinically, biologically and epidemiologically diverse, identified ~50 genes with differential distribution among them (Downing et al., 2011; Peacock et al., 2007). Another example is genome sequencing of 16 *L. donovani* strains which are responsible for visceral leishmaniasis in humans. These clinical isolates differ in their response to drug susceptibility and were obtained from infected patients in Nepal. Besides the expected relative genetic homogeneity, extensive variation in chromosome copy number between strains was found (Downing et al., 2011). Comparative genomic analyses unveiled genes such as those that are required by *L. donovani* to infect visceral organs (Peacock et al., 2007).

As it seems unlikely that any additional extensive and thorough genome analyzes of the dixenous genera *Leishmania* and *Trypanosoma* will reveal specific genetic elements responsible for successful parasite invasion of vertebrates, scientists are trying novel

approaches of deriving this information from the genomes of their sister groups, represented by the monoxenous insect trypanosomatids. This is a recent approach, as these medically and economically unimportant parasites were until few years ago represented by just a single partially annotated draft genome of *Crithidia fasciculata* with N50 scaffold size of 920 kb (the scaffold size above which 50% of the total length of the sequence assembly can be found) that was deposited in TriTrypDB database for public use (Aslett et al., 2010). Thanks to a number of recent studies, extensive sampling of insect trypanosomatids and constant drop of price per sequenced nucleotide base pair, ten genome drafts and numerous transcriptomes are now available for different lineages of monoxenous members of the family Trypanosomatidae. These genome draft assemblies are from *Lotmaria passim* (formerly known as *Crithidia mellificae*) a honey bee parasite with scaffold N50 = 32 kb (Runckel et al., 2014), *Leptomonas seymouri* (N50 = 70.6 kb) found in coinfections with *Leishmania* (Kraeva et al., 2015), *Blechomonas ayalai* (N50 = 94,8 kb) from flies and *Angomonas deanei* (N50 = 2.5 kb) and *Strigomonas culicis* (N50 = 2.7 kb), both endosymbiont-bearing species isolated from mosquitos and hemipterid insects (Motta et al., 2013). Furthermore, contigs of *Herpetomonas muscarum* (N50 = 6.8kb) have been deposited to GenBank (Alves et al., 2013).

The most notable example of a monoxenous genome is that of *Leptomonas pyrrhocoris* with a draft of 30 Mb, which was assembled almost to the chromosomal level and thus can serve as a benchmark for this group of trypanosomatids (Flegontov et al., 2016). Another noteworthy example is the first genome draft of the free-living kinetoplastid *Bodo saltans* with genome size of 40 Mb (Jackson et al., 2016). Except the transcriptome of *Neobodo designis* produced in frame of the Marine Microbial Eukaryote Transcriptome Sequencing Project (http://marinemicroeukaryotes.org), no other sequencing data are available from free-living bodonids.

Lately, we have sequenced the genome of *Paratrypanosoma confusum*, which constitutes the earliest branch between free-living bodonids and parasitic trypanosomatids. Its genome size is 31.6 Mb, being in term of quality on par with the genome of *L. pyrrhocoris* (discussed in more detail in chapter 3.2).

1.2.5 RNA interference

A key approach to study gene function in trypanosomatids rests in employing RNA interference. This machinery is responsible for maintaining genome stability by silencing mobile elements and repeats (Patrick et al., 2008), has a role in heterochromatin formation (Durand-Dubief et al., 2007), provides defense against invading viruses (Lye et al., 2010) and regulates mRNA levels, thus gene expression by specific destruction of mRNA molecules. Most detailed information about the RNAi pathway in protist parasites is available from *T. brucei,* where it has two branches. One is located in the nucleus where it is initiated by a Dicer-like enzyme *Tb*DCL2, which digests long RNA duplexes (dsRNA) into both 20-30 nt-long siRNA duplexes and 40-50 nt-long dsRNA intermediates that are being provided to the second RNAi arm represented by *Tb*DCL1 enzyme, which is located in the cytoplasm. This enzyme further processes dsRNA into siRNA duplexes. The nuclear part is believed to be mainly responsible for the down-regulation of retroposons and repeats and thus maintaining genome stability, while the cytosolic one is considered to be a secondary checkpoint against retroposons and repeats that escaped from the nucleus (Patrick et al., 2009). Working together, these enzymes enable full RNAi response, each providing a "program" for Argonaute AGO1. The siRNA duplexes are separated into single-stranded siRNA by DnaQ family 3´-5´exonuclease *Tb*RIF4 and possible *Tb*DCL1 cofactor, *Tb*RIF5 (Barnes et al., 2012). AGO1 loading with these single-stranded siRNA is done by the Hsp90/70 chaperone system (Iwasaki et al., 2010), and terminal 3´ ribose is modified with HEN1 methyltransferase, thus protecting it against nucleases. The whole complex forms the RNA-induced silencing complex that seeks and cleaves homologous target RNA.

The discovery of a functional RNAi pathway in *T. brucei* (Ngo et al., 1998) and its effectiveness and usefulness in functional studies provoked its search in other kinetoplastids. It was later found to be functional in other African trypanosomes, *Crithidia fasciculata*, *Leishmania braziliensis,* and *Leishmania guyanensis*, albeit with low efficiency (Lye et al., 2010). Unfortunately, *T. cruzi* (DaRocha et al., 2004), *Leishmania major*, *Leishmania donovani* (Robinson and Beverley, 2003), *Leishmania infantum* or *Leptomonas seymouri* possess core components of the RNAi pathway

The finding of a highly efficient RNAi pathway in *L. brasiliensis* was surprising and triggered speculations as to why the pathway was lost in the Old World leishmanias, but retained in their New World kins. In one scenario, it was proposed that the forces leading to the loss of RNAi pathway in protist parasites are dsRNA viruses termed LRV (Lye et al.,

2010). It seems that it is beneficial for *Leishmania* species to harbor dsRNA viruses, as upon release they are capable of modulating the vertebrate host immune system, thus increasing the survival (Gupta and Deep, 2007) and pathogenicity (Ives et al., 2011) of the flagellate.

Recent genome sequencing of bodonids and insect trypanosomatids revealed that the RNAi pathway is present and probably functional in several other kinetoplastids, such as *Leptomonas pyrrhocoris* (Flegontov et al., 2016), *Bodo saltans, T. congolense, T. evansi, T. grayi, T. vivax* and in the most basal branching parasitic trypanosomatid *Paratrypanosoma confusum* (Tab. 1). Moreover, *P. confusum* harbors dsRNA virus particles from the *Narnavirus* family (D. Grybchuk, unpublished results). These finding are parsimonious with a scenario, according to which RNAi machinery was lost only later in the evolution of trypanosomatid parasites.

Table 1: Presence/absence of RNAi pathway core genes in 30 analyzed species. Presence is marked with green color. Numbers represent number of gene copies in each species.

Gene name	DCL1	DCL2	AGO1	RIF4	RIF5
B. ayalai	0	1	0	0	0
B. saltans	2	2	1	4	0
C. fasciculata	1	1	1	1	1
L. braziliensis M2903	1	1	1	1	1
L. braziliensis M2904	2	1	1	1	1
L. donovani	1	1	1	0	0
L. infantum	1	1	1	0	0
L. major Friedlin	1	1	1	0	0
L. mexicana	1	1	1	0	0
L. tarentolae	1	1	1	1	1
L. pyrrhocoris	3	1	2	1	3
L. seymouri	0	1	0	0	0
N. designis	0	1	0	0	0
N. gruberi	0	0	0	0	0
P. confusum	1	1	1	1	0
Phytomonas sp. EM1	0	2	0	0	0
Perkinsela sp. CCAP-1560	1	1	0	0	0
Phytomonas sp. HART1	0	2	0	0	0
T. borreli	0	1	1	1	0
T. brucei Lister 427	1	1	1	1	1
T. brucei gambiense	1	1	1	1	1
T. brucei TREU927	1	1	1	1	1
T. congolense	2	1	1	1	2
T. cruzi Esmeraldo-like	0	0	0	0	0
T. cruzi Non-Esmeraldo	0	1	0	0	0
T. cruzi Sylvio	0	1	0	0	0
T. evansi	1	1	1	1	1
T. grayi	1	1	1	1	1
T. cruzi marinkellei	0	1	0	0	0
T. vivax	1	1	1	1	1

2 Research objectives

- Investigation of the diversity and host specificity of trypanosomatids
- Investigation of morphology and biology of monoxenous trypanosomatid *P. confusum*
- Genome and transcriptome sequencing of *P. confusum*
- Genome assembly and annotation of *P. confusum*
- Use comparative genomic approaches to shed light on trypanosomatid biology and evolution

3 Summary of results and discussion

Highlights of the core findings of the presented thesis:

(i) Tsetse flies (*Glossina* spp.) transmit several new, previously unknown *Trypanosoma* species from the *Trypanosoma simiae* and *Trypanosoma congolense* complexes. *Trypanosoma* sp. 'Msubugwe' that we have identified forms an enigmatic group that possibly represents an evolutionary link between the subgenera *Trypanozoon* and *Nannomonas.* High-throughput analysis of DNA isolated from blood meal from tsetse flies proved that they mainly feed on wild ruminants, humans and suids.

(ii) *P. confusum* is the earliest branching insect trypanosomatid constituting basal branch on transition between free-living bodonids and parasitic trypanosomatids.

(iii) *P. confusum* is morphologically very flexible, creating *in vitro* three distinct life stages that are common in monoxenous and dixenous trypanosomatids. This finding provides evidence that the ancestral trypanosomatid was already equipped with morphological flexibility that proved very suitable during the evolution of parasitism. Moreover, *P. confusum* retains cytostome, a feature known so far only in *T. cruzi.*

(iv) While *P. confusum* has already diverged from the common trypanosomatid ancestor, it still possesses more ancestral features than other trypanosomatid lineages except the stercorarian trypanosomes.

3.1 Survey of new trypanosomes transmitted by tsetse fly

The primary goal of the first project of this thesis was to broaden our knowledge of trypanosomes transmitted by tsetse and tabanid flies in Africa. Some trypanosomes are serious human and livestock parasites with high medical and economic impact. The most significant include five sub-species from the *Trypanosoma brucei* complex and *Trypanosoma congolense, Trypanosoma vivax* and *Trypanosoma simiae*. Trypanosomes are the most species-rich genus with a wide spectrum of infected hosts throughout the animal kingdom, and even though new species are being described regularly (Adams et al., 2010; Hamilton et al., 2008), it is evident that the majority of their real diversity remains hidden. One of the methods how to elucidate this diversity is collecting and sampling possible vectors. It is true that the presence of trypanosome DNA in bloodsucking flies cannot definitely answer whether or not they are being transmitted to vertebrate hosts. Yet, their presence in extracted blood meal from gut of flies strongly indicates which trypanosomes infect these vertebrates (Muturi et al., 2011). Moreover, this approach may be also used as sampling tool for vertebrate diversity in the region of bloodsucking fly capture (Calvignac-Spencer et al., 2013). Therefore, we used the opportunity to collect samples in the wild habitat of African great apes – the Dzanga-Sangha Protected Areas, Central African Republic. To our knowledge, this was the first time instance of sampling tsetse and tabanid flies in the primary forest inhabited by African great apes.

In total, 1033 individual DNA samples originating from engorged tsetse flies (*G. tabaniformis* and *G. f. fuscipes*) were processed. Due to the low quantity of isolated DNA from each sample, we identified trypanosome species using the nested PCR approach based on the 18S rRNA gene. This approach identified eight trypanosome species falling into three subgenera, namely six species from the subgenus *Nannomonas,* two species from the subgenus *Trypanozoon* and one *Megatrypanum* species. The two newly identified members of the subgenus *Trypanozoon* are particularly interesting, forming a group that may represent a link between the subgenera *Trypanozoon* and *Nannomonas.*

To investigate the original vertebrate species of tsetse blood meals, we have used the advantages of high-throughput analysis based on library sequencing (454 Roche) using amplicons created from PCR-amplified mitochondrial cytochrome b (cytb) gene from all 1033 samples that proved as a good determining marker of sources of bloodmeal in tsetse flies (Muturi et al., 2011). This highly sensitive approach confirmed feeding opportunism of tsetse flies. The principal source of blood were mammals, such as forest buffaloes (*Syncerus*

caffer nanus), humans, bongos and sitatungas (*Tragelaphus* spp.), wild hogs (*Potamochoerus porcus* and *Hylochoerus meinertzhageni*) and duikers (*Cephalophus* spp.), while sequences from unspecified rodent or forest elephants (*Loxodonta cyclotis*) were identified scarcely. We were also able to show that tsetse flies occasionally feed on tortoises (Testudinidae) and crocodiles (*Osteolaemus tetraspis*). Even when lowland gorillas frequented the capturing sites of tsetse flies, it seems that they are not being targeted, as no sequences from these mammals were found.

3.2 *Paratrypanosoma confusum,* a unique trypanosomatid

The second part of the thesis focuses on the morphology and biology of *P. confusum* and its bioinformatic analysis, aiming to gain a greater understanding on its fascinating features and possibly also on the evolution of trypanosomatids. *P. confusum* represents the basal branch of obligatory parasitic trypanosomatids. After we had initially isolated and verified its early-branching position (Flegontov et al., 2013), we decided to investigate the interesting features of this species, because its phylogenetic position makes it particularly informative for tracing the evolution of obligatory parasitism.

Using light and electron microscopy (EM) we successfully demonstrated that *P. confusum* assumes strikingly different morphotypes. In the axenic culture, *Paratrypanosoma* alternates between two morphologically distinct stages: a highly motile promastigote equipped with a long flagellum (Fig. 4A), and a sedentary stage, somewhat similar to a haptomonad of *Leishmania*, but completely lacking an exogenous flagellum (Figs. 4A–E). When cultivated on semi-solid agar plates it creates a third stage, which resembles very much the amastigote of *Leishmania* species (Fig. 4A). EM also revealed that *Paratrypanosoma* possesses a cytostome, a structure known only from *T. cruzi* (Alcantara et al., 2014).

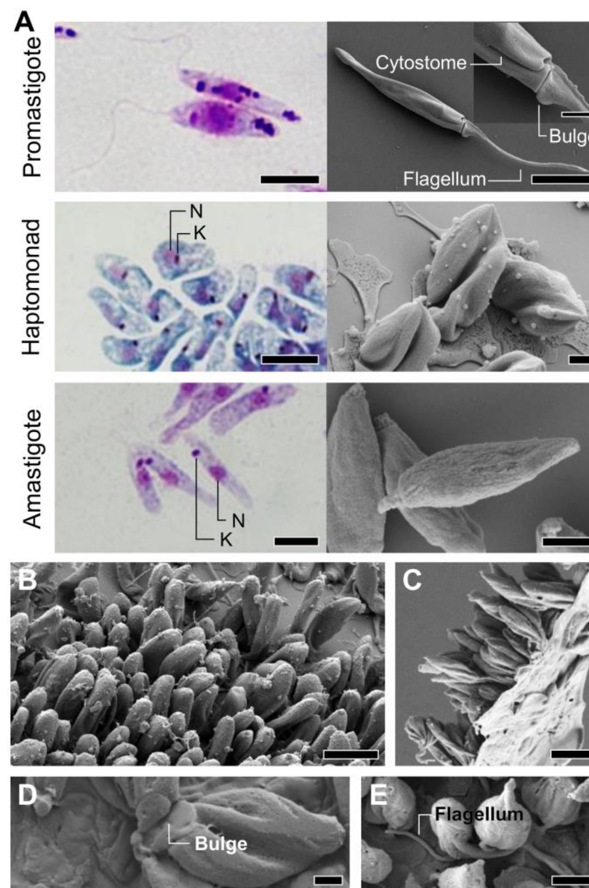


Figure 4. Morphologies of *Paratrypanosoma*. A) Different morphologies by light microscopy of Giemsa-stained cells (left) and SEM (right). Promastigote has a long flagellum with a prominent bulge at its base (inset). The kinetoplast (K) and nucleus (N) are indicated. Neither haptomonad nor amastigote have a long flagellum. **B-E)** SEM of haptomonads. **B)** Haptomonads attach perpendicular to the surface in dense clusters. **C)** Haptomonad cluster detached from the surface by propylene oxide showing the underside of the attachment pad. **D)** Haptomonads retain a bulge at the tip of their short flagellum. **E)** Occasional long flagella are visible extending from the haptomonad attachment pad. Scale bars: 5 µm (A, B, C; except in A middle image - 10 µm, and inset -1 µm); D) 1 µm; E) 2 µm.

The haptomonad stage forms an extensive thin attachment plaque on various surfaces (Figs. 4B-E). These morphologies have been observed in monoxenous trypanosomatids, however, it was almost impossible to study stages like the haptomonad in detail. It should be noted that these morphologies are tractable in this flagellate. Following the interstagial transformation by time-lapse video, antibodies recognizing proteins within the paraflagellar rod and EM of samples treated with propylene-oxide (Beattie and Gull, 1997) revealed that it initiates with the attachment of promastigote to the surface via a prominently extended basis of its external flagellum, termed here the "bulge" (Figs. 4A, D). This is followed by the shortening and disappearance of the external flagellum, likely by depolymerization, and repositioning of the now firmly attached cell into an upright position (Fig. 4A, C and Fig. 5A, B). At this stage, the external part of the flagellum disappears, the bulge becomes massively reshaped, as it transforms into a thin and amoeboid pad firmly attached to the

surface (Fig. 4A). A colony of haptomonads may cover the whole area with their extensive adhesive pads. Upon transformation into sessile haptomonads, the cells divide, producing dissimilar daughter cells, which either remain attached and divide again, or they can recreate the flagellum and thus regain motility (Figs. 5A, B). Similar observations with shortening of the flagellum or production of different daughter cells have been observed for *Leishmania* spp. and *T. brucei,* respectively (Wheeler et al., 2015; Wheeler et al., 2013).

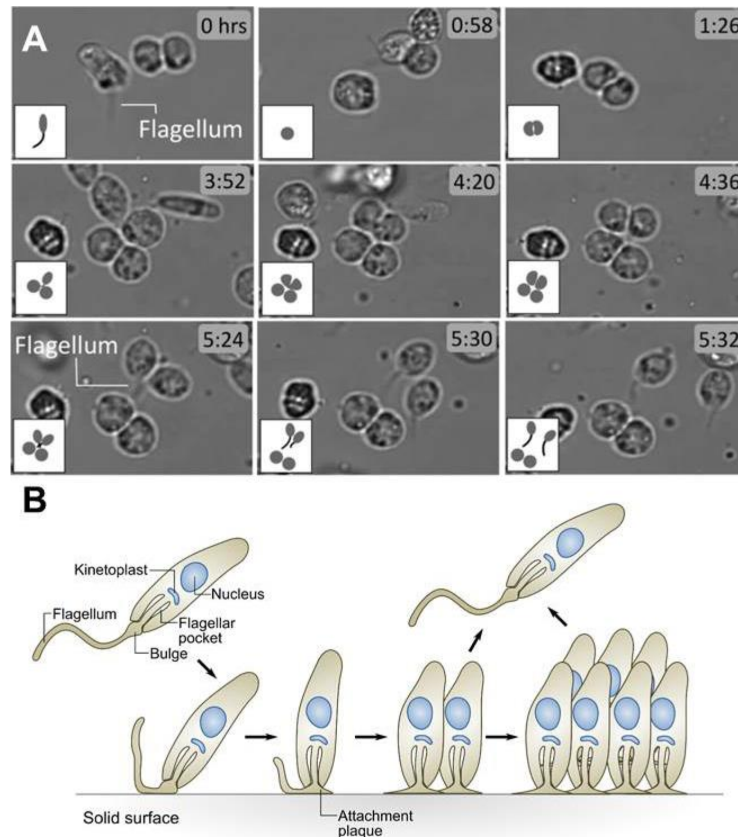


Figure 5. Promastigote attachment and replication as a haptomonad. A) Still images from time-lapse videos of promastigote and haptomonad interconversion. Attachment of a promastigote to form a haptomonad and two successive divisions; first forming two haptomonads, then two haptomonads and two promastigotes. Time (hours: minutes) for each frame is shown in the top right, and a cartoon of cell arrangement is shown in the bottom left. **B)** Cartoon of the attachment process. The promastigote freely swims and can attach to a surface by the bulge. The flagellum shortens and the cell assumes an upright position with the bulge expanding into a thin attachment pad. Division events can generate cells which either attach to the surface or grow a long flagellum and leave the colony.

This indicates that the ancestral trypanosomatid was likely endowed with an extreme morphological flexibility, which was amply used when it faced dramatically different conditions after invading a wide variety of (in)vertebrate hosts. Hence, the capacity for extensive interstagial transformations of *Leishmania* and *Trypanosoma* spp. during their life cycles that involves both the mammalian and insect hosts seem to have existed already in their single-host predecessor. Consequently, the wide array of trypanosomatid morphotypes apparently did not originate within the context of dixenous parasitism, but represents yet

another character state that predated the two-host life style (Janouskovec and Keeling, 2016). The ability to firmly but transiently attach to various surfaces with their dexterous flagellum likely equipped them with means preventing their discharge from the host. This has already been described in some monoxenous (e.g. *Leptomonas* spp.) and dixenous trypanosomatids (e.g. *Trypanosoma* and *Leishmania* spp.) (Gadelha et al., 2005), as well as in bodonids (Current, 1980).

As we were interested what is happening during this transformation on molecular level, we have sequenced the *Paratrypanosoma* genome as a reference for transcriptome analysis of different morphological stages. We obtained an assembly of N50 = 438 kb and annotated 8668 protein coding genes, 66 tRNAs and 122 copies of rRNA genes. The genome of *Paratrypanosoma* also encodes 8 of 9 meiosis-specific genes and 17 of 23 meiosis-associated genes, including two homologs required for gamete membrane fusion and karyogamy. According to BUSCO benchmarking (Simao et al., 2015), the quality of the assembly is on par with other high-quality trypanosomatid assemblies, containing 72.3% of core eukaryotic genes (for comparison, it is 74.9% in *T. brucei*, 73.6% in *L. major* and 72.6% in *L. pyrrhocoris*). Evaluation of differentially expressed genes revealed 324 and 267 significantly (fold change ≥ 2 and FDR-corrected p-value ≤ 0.05) overexpressed genes in promastigotes and haptomonads, respectively. For the latter, the most frequent and specific gene onthology (GO) terms are associated with ribosome biogenesis (ribosome, structural constituent of ribosome, translation, ribosome biogenesis and nucleolus), whereas for promastigotes, these are intermediary metabolism (oxidation-reduction processes, glycolysis and malate metabolic processes). These are not surprising results, when considering our observations that haptomonads divide massively and promastigotes need plenty of energy for the beating of the flagellum. However, only 48 and 124 differentially expressed gene in promastigotes and haptomonads, respectively, have homologues with assigned function, with the rest being hypothetical without either blast hits or functional annotations.

We have utilized comparative genomics to investigate if *Paratrypanosoma* displays more ancestral features than other trypanosomatids. Using the OrthoFinder (Emms and Kelly, 2015), we generated orthologous groups (OGs) of proteins for a set of trypanosomatids, bodonids (free-living *Bodo saltans* and *Neobodo designis*, a parasite *Trypanoplasma borreli*, and an endosymbiont *Perkinsela* sp.) and for *Naegleria gruberi*, a heterolobosean (Fig. 6A). Next, we counted the number of ancestral OGs in each species, i.e. those shared with any bodonid or *Naegleria*. *Trypanosoma grayi*, a parasite of crocodiles that belongs to the clade of stercorarian trypanosomes, shows the highest number of ancestral

OGs (6197), while *Paratrypanosoma* shows the second highest number (6066). Given that both species have the shortest branches in a multigene tree (Fig. 6A), this result is not unexpected. Thus, it is likely that these two species are the slowest-evolving trypanosomatids in the analyzed dataset.

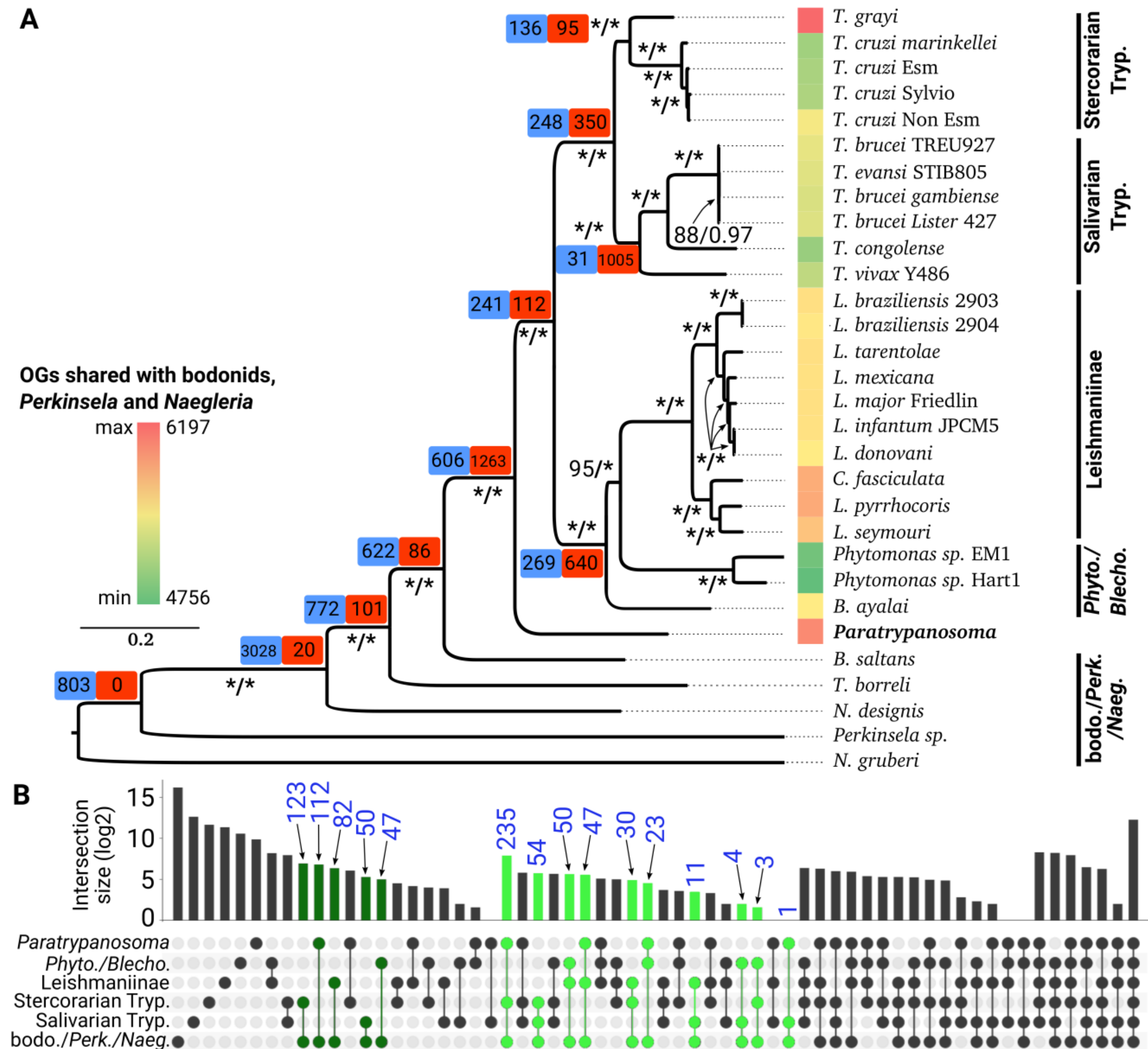


Figure 6. Ancestral gene families in *Paratrypanosoma* and other trypanosomatids. A) Phylogenetic tree based on a concatenated set of 98 proteins, constructed with maximum likelihood and Bayesian approaches. For each species, the counts of orthologous groups shared with at least one bodonid or *Naegleria* are color-coded (see a scale on the left). Support values for tree nodes are shown in the format 'bootstrap support / posterior probability', and full support is indicated with asterisks. Gains (in blue) and losses (in red) of OGs were mapped on the tree using the Dollo parsimony algorithm. Few basal nodes are shown for clarity. **B)** Phyletic patterns visualized for six clades: *Paratrypanosoma*, *Phytomonas/Blechomonas,* Leishmaniinae (*Leishmania, Leptomonas, Crithidia*), stercorarian trypanosomes, salivarian trypanosomes, and bodonids/*Perkinsela/Naegleria*. Counts of OGs unique for each clade and all possible intersections of the six sets are shown in the bar plot using a log scale (see selected OG counts above the bars). Counts of ancestral gene families shared by one or two trypanosomatid clades are highlighted in dark-green and light-green, respectively.

Tapping into the distribution of ancestral OGs throughout major groups (i/ *Paratrypanosoma*, ii/ salivarian (African trypanosomes) and iii/ stercorarian trypanosomes (American trypanosomes), iv/ Leishmaniinae (*Leishmania*, *Leptomonas*, *Crithidia*), v/ *Phytomonas*/*Blechomonas* and vi/ bodonids/*Perkinsela*/*Naegleria*) revealed that stercorarian trypanosomes and *Paratrypanosoma* have the highest numbers of ancestral OGs unique to those clades, 123 and 112, respectively (Fig. 6B). When we investigated those shared ancestral genes, and annotated their possible function, we found that even when most of those genes have been annotated as hypothetical proteins, some do stand out. These are genes of tryptophan and histidine catabolism, and arginine biosynthesis that are unique to *Paratrypanosoma* or the stercorarian trypanosomes, and were lost in the other trypanosomatids. The fifth largest protein family in *T. cruzi,* the dispersed gene family 1 (DFG1) (Lander et al., 2010) is unique to *Paratrypanosoma* and the stercorarian trypanosomes. DFG1 is a family of long membrane proteins of unknown function that possess glycosylated extracellular domain stored in intracellular vesicles that are excreted during the trypomastigote to amastigote transformation (Lacomble et al., 2009; Lander et al., 2010).

Recent comparison of the *B. saltans* and trypanosomatid genomes revealed that metabolic losses preceded parasitism (Jackson et al., 2016; Opperdoes et al., 2016), and no further gene loss or streamlining of the genome occurred in the crown trypanosomatids (Jackson et al., 2016). Our comparative analysis of OGs gains and losses confirm these results and while *P. confusum* has already diverged from the common ancestor, together with the stercorarian trypanosomes, it still retains numerous ancestral genes that have been lost in all other lineages. Unfortunately, most of these genes have been annotated with only hypothetical function, thus functional studies are required to determine their function.

4 REFERENCES

Adams, E. R., Hamilton, P. B., Rodrigues, A. C., Malele, II, Delespaux, V., Teixeira, M. M. and Gibson, W. (2010). New *Trypanosoma* (Duttonella) *vivax* genotypes from tsetse flies in East Africa. *Parasitology* **137**, 641-650.

Adl, S. M., Simpson, A. G., Lane, C. E., Lukeš, J., Bass, D., Bowser, S. S., Brown, M. W., Burki, F., Dunthorn, M., Hampl, V., Heiss, A., Hoppenrath, M., Lara, E., Le Gall, L., Lynn, D. H., McManus, H., Mitchell, E. A., Mozley-Stanridge, S. E., Parfrey, L. W., Pawlowski, J., Rueckert, S., Shadwick, L., Schoch, C. L., Smirnov, A., Spiegel, F. W. (2012). The revised classification of eukaryotes. *The J Eukaryot Microbiol* **59**, 429-493.

Alcantara, C. L., Vidal, J. C., de Souza, W. and Cunha-e-Silva, N. L. (2014). The three-dimensional structure of the cytostome-cytopharynx complex of *Trypanosoma cruzi* epimastigotes. *J Cell Sci* **127**, 2227-2237.

Alvar, J., Velez, I. D., Bern, C., Herrero, M., Desjeux, P., Cano, J., Jannin, J., den Boer, M. and Team, W. H. O. L. C. (2012). Leishmaniasis worldwide and global estimates of its incidence. *PloS One* **7**, e35671.

Alves, J. M., Klein, C. C., da Silva, F. M., Costa-Martins, A. G., Serrano, M. G., Buck, G. A., Vasconcelos, A. T., Sagot, M. F., Teixeira, M. M., Motta, M. C., Camargo, E. P. (2013). Endosymbiosis in trypanosomatids: the genomic cooperation between bacterium and host in the synthesis of essential amino acids is heavily influenced by multiple horizontal gene transfers. *BMC Evol Biol* **13**, 190.

Aslett, M., Aurrecoechea, C., Berriman, M., Brestelli, J., Brunk, B. P., Carrington, M., Depledge, D. P., Fischer, S., Gajria, B., Gao, X., Gardner, M. J.

Gingle, A., Grant, G., Harb, O. S., Heiges, M., Hertz-Fowler, C., Houston, R., Innamorato, F., Iodice, J., Kissinger, J. C., Kraemer, E., Li, W., Logan, F. J., Miller, J. A., Mitra, S., Myler, P. J., Nayak, V., Pennington, C., Phan, I., Pinney, D. F., Ramasamy, G., Rogers, M. B., Roos, D. S., Ross, C., Sivam, D., Smith, D. F., Srinivasamoorthy, G., Stoeckert, C. J., Jr., Subramanian, S., Thibodeau, R., Tivey, A., Treatman, C., Velarde, G., Wang, H. (2010). TriTrypDB: a functional genomic resource for the Trypanosomatidae. *Nucleic Acids Res* **38**, D457-462.

Banuls, A. L., Hide, M. and Prugnolle, F. (2007). *Leishmania* and the leishmaniases: a parasite genetic update and advances in taxonomy, epidemiology and pathogenicity in humans. *Adv Parasitol* **64**, 1-109.

Barnes, R. L., Shi, H., Kolev, N. G., Tschudi, C. and Ullu, E. (2012). Comparative genomics reveals two novel RNAi factors in *Trypanosoma brucei* and provides insight into the core machinery. *PLoS Pathog* **8**, e1002678.

Barratt, J., Kaufer, A., Peters, B., Craig, D., Lawrence, A., Roberts, T., Lee, R., McAuliffe, G., Stark, D. and Ellis, J. (2017). Isolation of Novel Trypanosomatid, *Zelonia australiensis* sp. nov. (Kinetoplastida: Trypanosomatidae) Provides Support for a Gondwanan Origin of Dixenous Parasitism in the Leishmaniinae. *PLOS Neglect Trop D* **11**, e0005215.

Barratt, J. L., Harkness, J., Marriott, D., Ellis, J. T. and Stark, D. (2010). Importance of nonenteric protozoan infections in immunocompromised people. *Clin Microbiol Rev* **23**, 795-836.

Barrett, M. P., Burchmore, R. J., Stich, A., Lazzari, J. O., Frasch, A. C., Cazzulo, J. J. and Krishna, S. (2003). The trypanosomiases. *Lancet* **362**, 1469-1480.

Batistoti, M., Cavazzana, M., Jr., Serrano, M. G., Ogatta, S. F., Baccan, G. C., Jankevicius, J. V., Teixeira, M. M. and Jankevicius, S. I. (2001). Genetic variability of trypanosomatids isolated from phytophagous hemiptera defined by morphological, biochemical, and molecular taxonomic markers. *J Parasitol* **87**, 1335-1341.

Beattie, P. and Gull, K. (1997). Cytoskeletal architecture and components involved in the attachment of *Trypanosoma congolense* epimastigotes. *Parasitology* **115 (Pt 1)**, 47-55.

Brun, R., Hecker, H. and Lun, Z. R. (1998). Trypanosoma evansi and *T. equiperdum*: distribution, biology, treatment and phylogenetic relationship (a review). *Vet Parasitol* **79**, 95-107.

Butler, C. E., Jaskowska, E. and Kelly, S. (2017). Genome Sequence of *Phytomonas francai*, a Cassava (*Manihot esculenta*) Latex Parasite. *Genome Announc* **5**.

Calvignac-Spencer, S., Leendertz, F. H., Gilbert, M. T. and Schubert, G. (2013). An invertebrate stomach's view on vertebrate ecology: certain invertebrates could be used as "vertebrate samplers" and deliver DNA-based information on many aspects of vertebrate ecology. *BioEssays* **35**, 1004-1013.

Camargo, E. P. (1999). *Phytomonas* and other trypanosomatid parasites of plants and fruit. *Adv Parasitol* **42**, 29-112.

Campbell, D. A., Sturm, N. R. and Yu, M. C. (2000). Transcription of the kinetoplastid spliced leader RNA gene. *Parasitol Today* **16**, 78-82.

Cavalier-Smith, T. (2010). Kingdoms Protozoa and Chromista and the eozoan root of the eukaryotic tree. *Biol Lett* **6**, 342-345.

Clayton, C. E. (2016). Gene expression in Kinetoplastids. *Curr Opin Microbiol* **32**, 46-51.

Cupolillo, E., Medina-Acosta, E., Noyes, H., Momen, H. and Grimaldi, G., Jr. (2000). A revised classification for *Leishmania* and *Endotrypanum*. *Parasitol Today* **16**, 142-144.

Current, W. L. (1980). *Cryptobia* sp. in the snail *Triadopsis multilineata* (Say): Fine structure of attached flagellates and their mode of attachment to the spermatheca. *J Protozool* **27**, 278-287.

d'Avila-Levy, C. M., Boucinha, C., Kostygov, A., Santos, H. L., Morelli, K. A., Grybchuk-Ieremenko, A., Duval, L., Votýpka, J., Yurchenko, V., Grellier, P., Lukeš, J. (2015). Exploring the environmental diversity of kinetoplastid flagellates in the high-throughput DNA sequencing era. *Mem I Oswaldo Cruz* **110**, 956-965.

DaRocha, W. D., Otsu, K., Teixeira, S. M. and Donelson, J. E. (2004). Tests of cytoplasmic RNA interference (RNAi) and construction of a tetracycline-inducible T7 promoter system in *Trypanosoma cruzi*. *Mol Biochem Parasit* **133**, 175-186.

David, V., Flegontov, P., Gerasimov, E., Tanifuji, G., Hashimi, H., Logacheva, M. D., Maruyama, S., Onodera, N. T., Gray, M. W., Archibald, J. M., Lukeš, J. (2015). Gene Loss and Error-Prone RNA Editing in the Mitochondrion of *Perkinsela*, an Endosymbiotic Kinetoplastid. *MBio* **6**, e01498-01415.

De Souza, W. (1995). Structural organization of the cell surface of pathogenic protozoa. *Micron* **26**, 405-430.

Downing, T., Imamura, H., Decuypere, S., Clark, T. G., Coombs, G. H., Cotton, J. A., Hilley, J. D., de Doncker, S., Maes, I., Mottram, J. C., Quail, M. A., Rijal, S., Sanders, M., Schonian, G., Stark, O., Sundar, S., Vanaerschot, M., Hertz-Fowler, C., Dujardin, J. C., Berriman, M. (2011). Whole genome sequencing of multiple *Leishmania donovani* clinical isolates provides insights into population structure and mechanisms of drug resistance. *Genome Res* **21**, 2143-2156.

Durand-Dubief, M., Absalon, S., Menzer, L., Ngwabyt, S., Ersfeld, K. and Bastin, P. (2007). The Argonaute protein TbAGO1 contributes to large and mini-chromosome segregation and is required for control of RIME retroposons and RHS pseudogene-associated transcripts. *Mol Biochem Parasit* **156**, 144-153.

Dyková, I., Fiala, I. and Peckova, H. (2008). *Neoparamoeba* spp. and their eukaryotic endosymbionts similar to Perkinsela amoebae (Hollande, 1980): coevolution demonstrated by SSU rRNA gene phylogenies. *Eur J Protistol* **44**, 269-277.

El-Sayed, N. M., Myler, P. J., Blandin, G., Berriman, M., Crabtree, J., Aggarwal, G., Caler, E., Renauld, H., Worthey, E. A., Hertz-Fowler, C., Ghedin, E., Peacock, C., Bartholomeu, D. C., Haas, B. J., Tran, A. N., Wortman, J. R., Alsmark, U. C., Angiuoli, S., Anupama, A., Badger, J., Bringaud, F., Cadag, E., Carlton, J. M., Cerqueira, G. C., Creasy, T., Delcher, A. L., Djikeng, A., Embley, T. M., Hauser, C., Ivens, A. C., Kummerfeld, S. K., Pereira-Leal, J. B., Nilsson, D., Peterson, J., Salzberg, S. L., Shallom, J., Silva, J. C., Sundaram, J., Westenberger, S., White, O., Melville, S. E., Donelson, J. E., Andersson, B., Stuart, K. D., Hall, N. (2005). Comparative genomics of trypanosomatid parasitic protozoa. *Science* **309**, 404-409.

Emms, D. M. and Kelly, S. (2015). OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* **16**, 157.

Espinosa, O. A., Serrano, M. G., Camargo, E. P., Teixeira, M. M. and Shaw, J. J. (2016). An appraisal of the taxonomy and nomenclature of trypanosomatids presently classified as Leishmania and Endotrypanum. *Parasitology*, 1-13.

Fadda, A., Ryten, M., Droll, D., Rojas, F., Farber, V., Haanstra, J. R., Merce, C., Bakker, B. M., Matthews, K. and Clayton, C. (2014). Transcriptome-wide analysis of trypanosome mRNA decay reveals complex degradation kinetics and suggests a role for co-transcriptional degradation in determining mRNA levels. *Mol Microbiol* **94**, 307-326.

Fermino, B. R., Viola, L. B., Paiva, F., Garcia, H. A., de Paula, C. D., Botero-Arias, R., Takata, C. S., Campaner, M., Hamilton, P. B., Camargo, E. P., Teixeira, M. M. (2013). 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* **6**, 313.

Flegontov, P., Butenko, A., Firsov, S., Kraeva, N., Eliáš, M., Field, M. C., Filatov, D., Flegontova, O., Gerasimov, E. S., Hlaváčová, J., Ishemgulova, A., Jackson, A. P., Kelly, S., Kostygov, A. Y., Logacheva, M. D., Maslov, D. A., Opperdoes, F. R., O'Reilly, A., Sadlova, J., Sevcikova, T., Venkatesh, D., Vlcek, C., Volf, P., Votypka, J., Zahonova, K., Yurchenko, V., Lukes, J. (2016). Genome of *Leptomonas pyrrhocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci Rep* **6**, 23704.

Flegontov, P., Votýpka, J., Skalický, T., Logacheva, M. D., Penin, A. A., Tanifuji, G., Onodera, N. T., Kondrashov, A. S., Volf, P., Archibald, J. M., Lukeš, J. (2013). *Paratrypanosoma* Is a Novel Early-Branching Trypanosomatid. *Curr Biol* **23**, 1787-1793.

Franco, J. R., Simarro, P. P., Diarra, A. and Jannin, J. G. (2014). Epidemiology of human African trypanosomiasis. *Clin Epidemiol* **6**, 257-275.

Gadelha, C., Wickstead, B., de Souza, W., Gull, K. and Cunha-e-Silva, N. (2005). Cryptic paraflagellar rod in endosymbiont-containing kinetoplastid protozoa. *Eukaryot Cell* **4**, 516-525.

Gibson, W. (2012). The origins of the trypanosome genome strains *Trypanosoma brucei brucei* TREU 927, T. b. gambiense DAL 972, T. vivax Y486 and T. congolense IL3000. *Parasit Vectors* **5**, 71.

Gluenz, E., Wheeler, R. J., Hughes, L. and Vaughan, S. (2015). Scanning and three-dimensional electron microscopy methods for the study of *Trypanosoma brucei* and *Leishmania mexicana* flagella. *Method Cell Biol* **127**, 509-542.

Greif, G., Ponce de Leon, M., Lamolle, G., Rodriguez, M., Pineyro, D., Tavares-Marques, L. M., Reyna-Bello, A., Robello, C. and Alvarez-Valin, F. (2013). Transcriptome analysis of the bloodstream stage from the parasite *Trypanosoma vivax*. *BMC genomics* **14**, 149.

Gupta, V. and Deep, A. (2007). An insight into the Leishmania RNA virus. Indian J Med Microbiol **25**, 7-9.

Hamilton, P. B., Adams, E. R., Malele, II and Gibson, W. C. (2008). A novel, high-throughput technique for species identification reveals a new species of tsetse-transmitted trypanosome related to the *Trypanosoma brucei* subgenus, *Trypanozoon*. *Infect Genet Evol* **8**, 26-33.

Hamilton, P. B., Stevens, J. R., Gaunt, M. W., Gidley, J. and Gibson, W. C. (2004). Trypanosomes are monophyletic: evidence from genes for glyceraldehyde phosphate dehydrogenase and small subunit ribosomal RNA. *Int J Parasitol* **34**, 1393-1404.

Hamilton, P. T., Votýpka, J., Dostalová, A., Yurchenko, V., Bird, N. H., Lukeš, J., Lemaitre, B. and Perlman, S. J. (2015). Infection Dynamics and Immune Response in a Newly Described Drosophila-Trypanosomatid Association. *MBio* **6**, e01356-01315.

Harkins, K. M., Schwartz, R. S., Cartwright, R. A. and Stone, A. C. (2016). Phylogenomic reconstruction supports supercontinent origins for Leishmania. *Infect Genet Evol***38**, 101-109.

Hayes, P. M., Lawton, S. P., Smit, N. J., Gibson, W. C. and Davies, A. J. (2014). Morphological and molecular characterization of a marine fish trypanosome from South Africa, including its development in a leech vector. *Parasit Vectors* **7**, 50.

Hemphill, A., Lawson, D. and Seebeck, T. (1991). The cytoskeletal architecture of *Trypanosoma brucei*. *J Parasitol* **77**, 603-612.

Hoare, C. A. and Wallace, F. G. (1966). Developmental stages of trypanosomatid flagellates: a new terminology. *Nature* **212**, 1385-1386.

Hollar, L., Lukeš, J. and Maslov, D. A. (1998). Monophyly of endosymbiont containing trypanosomatids: phylogeny versus taxonomy. *The J Eukaryot Microbiol* **45**, 293-297.

Chicharro, C. and Alvar, J. (2003). Lower trypanosomatids in HIV/AIDS patients. *Ann Trop Med Parasitol* **97 Suppl 1**, 75-78.

Iwasaki, S., Kobayashi, M., Yoda, M., Sakaguchi, Y., Katsuma, S., Suzuki, T. and Tomari, Y. (2010). Hsc70/Hsp90 chaperone machinery mediates ATP-dependent RISC loading of small RNA duplexes. *Mol Cell* **39**, 292-299.

Ives, A., Ronet, C., Prevel, F., Ruzzante, G., Fuertes-Marraco, S., Schutz, F., Zangger, H., Revaz-Breton, M., Lye, L. F., Hickerson, S. M., Beverley, S. M., Acha-Orbea, H., Launois, P., Fasel, N., Masina, S. (2011). Leishmania RNA virus controls the severity of mucocutaneous leishmaniasis. Science **331**, 775-778.

Jackson, A. P. (2015). Genome evolution in trypanosomatid parasites. *Parasitology* **142 Suppl 1**, S40-56.

Jackson, A. P., Otto, T. D., Aslett, M., Armstrong, S. D., Bringaud, F., Schlacht, A., Hartley, C., Sanders, M., Wastling, J. M., Dacks, J. B., Acosta-Serrano, A., Field, M. C., Ginger, M. L., Berriman, M. (2016). Kinetoplastid Phylogenomics Reveals the Evolutionary Innovations Associated with the Origins of Parasitism. *Curr Biol* **26**, 161-172.

Jackson, Y., Pinto, A. and Pett, S. (2014). Chagas disease in Australia and New Zealand: risks and needs for public health interventions. *Trop Med Int Health* **19**, 212-218.

Janouškovec, J. and Keeling, P. J. (2016). Evolution: Causality and the Origin of Parasitism. *Curr Biol* **26**, R174-177.

Jaskowska, E., Butler, C., Preston, G. and Kelly, S. (2015). *Phytomonas*: trypanosomatids adapted to plant environments. *PLoS Pathog* **11**, e1004484.

Jirků, M., Yurchenko, V. Y., Lukeš, J. and Maslov, D. A. (2012). New species of insect trypanosomatids from Costa Rica and the proposal for a new subfamily within the Trypanosomatidae. *The J Eukaryot Microbiol* **59**, 537-547.

Kaufer, A., Ellis, J., Stark, D. and Barratt, J. (2017). The evolution of trypanosomatid taxonomy. *Parasit Vectors* **10**, 287.

Kelly, S., Kramer, S., Schwede, A., Maini, P. K., Gull, K. and Carrington, M. (2012). Genome organization is a major component of gene expression control in response to stress and during the cell division cycle in trypanosomes. *Open Biol* **2**, 120033.

Kostygov, A., Malysheva, M. N. and Frolov, A. O. (2011). [Investigation of causes of the conflict between taxonomy and molecular phylogeny of trypanosomatids by the example of Leptomonas nabiculae podlipaev, 1987]. *Parazitologiia* **45**, 409-424.

Kostygov, A. and Yurchenko, V. (in press). Revised classification of the subfamily Leishmaniinae (Kinetoplastida). *Folia Parasit*.

Kostygov, A. Y., Dobáková, E., Grybchuk-Ieremenko, A., Váhala, D., Maslov, D. A., Votýpka, J., Lukeš, J. and Yurchenko, V. (2016). Novel Trypanosomatid-Bacterium Association: Evolution of Endosymbiosis in Action. *MBio* **7**, e01985.

Kostygov, A. Y., Grybchuk-Ieremenko, A., Malysheva, M. N., Frolov, A. O. and Yurchenko, V. (2014). Molecular revision of the genus *Wallaceina*. *Protist* **165**, 594-604.

Kozminsky, E., Kraeva, N., Ishemgulova, A., Dobáková, E., Lukeš, J., Kment, P., Yurchenko, V., Votýpka, J. and Maslov, D. A. (2015). Host-specificity of Monoxenous Trypanosomatids: Statistical Analysis of the Distribution and Transmission Patterns of the Parasites from Neotropical Heteroptera. *Protist* **166**, 551-568.

Kraeva, N., Butenko, A., Hlaváčová, J., Kostygov, A., Myškova, J., Grybchuk, D., Leštinová, T., Votýpka, J., Volf, P., Opperdoes, F., Flegontov, P., Lukes, J., Yurchenko, V. (2015). Leptomonas seymouri: Adaptations to the Dixenous Life Cycle Analyzed by Genome Sequencing, Transcriptome Profiling and Co-infection with *Leishmania donovani*. *PLoS Pathog* **11**, e1005127.

Kumagai, A., Ito, H. and Sasaki, R. (2013). Detection of the kinetoplastid *Azumiobodo hoyamushi*, the causative agent of soft tunic syndrome, in wild ascidians *Halocynthia roretzi*. *Dis Aquat Organ* **106**, 267-271.

Lacomble, S., Portman, N. and Gull, K. (2009). A protein-protein interaction map of the *Trypanosoma brucei* paraflagellar rod. *PloS One* **4**, e7685.

Lai, D. H., Hashimi, H., Lun, Z. R., Ayala, F. J. and Lukeš, J. (2008). Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T. brucei*. *P Natl Acad Sci USA* **105**, 1999-2004.

Lander, N., Bernal, C., Diez, N., Anez, N., Docampo, R. and Ramirez, J. L. (2010). Localization and developmental regulation of a dispersed gene family 1 protein in *Trypanosoma cruzi*. *Infect Immun* **78**, 231-240.

LeBowitz, J. H., Smith, H. Q., Rusche, L. and Beverley, S. M. (1993). Coupling of poly(A) site selection and trans-splicing in *Leishmania*. *Genes Dev* **7**, 996-1007.

Léger, L. (1904). Surles affinites de l'Herpetomonas subulata et la phylogene des trypanosomes. *Comp. Rend. Seances Soc. Biol. Ses. Fil.* **56**, 615-617.

Leonard, G., Soanes, D. M. and Stevens, J. R. (2011). Resolving the question of trypanosome monophyly: a comparative genomics approach using whole genome data sets with low taxon sampling. *Infect Genet Evol* **11**, 955-959.

Liu, B., Liu, Y., Motyka, S. A., Agbo, E. E. and Englund, P. T. (2005). Fellowship of the rings: the replication of kinetoplast DNA. *Trends Parasitol* **21**, 363-369.

Lukeš, J., Arts, G. J., van den Burg, J., de Haan, A., Opperdoes, F., Sloof, P. and Benne, R. (1994). Novel pattern of editing regions in mitochondrial transcripts of the cryptobiid *Trypanoplasma borreli*. *EMBO J* **13**, 5086-5098.

Lukeš, J., Mauricio, I. L., Schonian, G., Dujardin, J. C., Soteriadou, K., Dedet, J. P., Kuhls, K., Tintaya, K. W., Jirků, M., Chocholová, E., Haralambous, C., Pratlong, F., Obornik, M., Horak, A., Ayala, F. J., Miles, M. A. (2007). Evolutionary and geographical history of the *Leishmania donovani* complex with a revision of current taxonomy. *P Natl Acad Sci USA* **104**, 9375-9380.

Lukeš, J., Skalický, T., Týč, J., Votýpka, J. and Yurchenko, V. (2014). Evolution of parasitism in kinetoplastid flagellates. *Mol Biochem Parasit* **195**, 115-122.

Lye, L. F., Owens, K., Shi, H., Murta, S. M., Vieira, A. C., Turco, S. J., Tschudi, C., Ullu, E. and Beverley, S. M. (2010). Retention and loss of RNA interference pathways in trypanosomatid protozoans. *PLoS Pathog* **6**, e1001161.

Mair, G., Shi, H., Li, H., Djikeng, A., Aviles, H. O., Bishop, J. R., Falcone, F. H., Gavrilescu, C., Montgomery, J. L., Santori, M. I., Stern, L. S., Wang, Z., Ullu, E., Tschudi, C. (2000). A new twist in trypanosome RNA metabolism: cis-splicing of pre-mRNA. *RNA* **6**, 163-169.

Martinez-Calvillo, S., Yan, S., Nguyen, D., Fox, M., Stuart, K. and Myler, P. J. (2003). Transcription of *Leishmania major* Friedlin chromosome 1 initiates in both directions within a single region. *Mol Cell* **11**, 1291-1299.

Maslov, D. A., Lukeš, J., Jirků, M. and Simpson, L. (1996). Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. *Mol Biochem Parasit* **75**, 197-205.

Maslov, D. A., Votýpka, J., Yurchenko, V. and Lukeš, J. (2013). Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends Parasitol* **29**, 43-52.

Maslov, D. A., Yurchenko, V. Y., Jirků, M. and Lukeš, J. (2010). Two new species of trypanosomatid parasites isolated from *Heteroptera* in Costa Rica. *The J Eukaryot Microbiol* **57**, 177-188.

Mayer, M. G. and Floeter-Winter, L. M. (2005). Pre-mRNA trans-splicing: from kinetoplastids to mammals, an easy language for life diversity. *Mem I Oswaldo Cruz* **100**, 501-513.

Mesnil, F. and Brimont, E. (1908). Sur un hématozoaire nouveau (Endotrypanum n. gen.) d'un édenté de la Guyane. *C.R. Séances Soc. Biol. Ses. Fil.* **65**, 581-583.

Minchin, E. A. (1908). Investigations on the development of trypanosomes in the tsetse flies and other Diptera. *Q. J. Microsc. Sci.* **52**, 159-260.

Morsy, T. A., Schnur, L. F., Feinsod, F. M., Michael, S. A., Saah, A., Salama, M. M. and Wahba, M. M. (1988). The discovery and preliminary characterization of a novel trypanosomatid

parasite from *Rattus norvegicus* and stray dogs from Alexandria, Egypt. *Ann Trop Med Parasitol* **82**, 437-444.

Motta, M. C., Martins, A. C., de Souza, S. S., Catta-Preta, C. M., Silva, R., Klein, C. C., de Almeida, L. G., de Lima Cunha, O., Ciapina, L. P., Brocchi, M., Colabardini, A. C., de Araujo Lima, B., Machado, C. R., de Almeida Soares, C. M., Probst, C. M., de Menezes, C. B., Thompson, C. E., Bartholomeu, D. C., Gradia, D. F., Pavoni, D. P., Grisard, E. C., Fantinatti-Garboggini, F., Marchini, F. K., Rodrigues-Luiz, G. F., Wagner, G., Goldman, G. H., Fietto, J. L., Elias, M. C., Goldman, M. H., Sagot, M. F., Pereira, M., Stoco, P. H., de Mendonca-Neto, R. P., Teixeira, S. M., Maciel, T. E., de Oliveira Mendes, T. A., Urmenyi, T. P., de Souza, W., Schenkman, S., de Vasconcelos, A. T. (2013). Predicting the proteins of *Angomonas deanei*, *Strigomonas culicis* and their respective endosymbionts reveals new aspects of the trypanosomatidae family. *PloS one* **8**, e60209.

Muturi, C. N., Ouma, J. O., Malele, II, Ngure, R. M., Rutto, J. J., Mithofer, K. M., Enyaru, J. and Masiga, D. K. (2011). Tracking the feeding patterns of tsetse flies (*Glossina genus*) by analysis of bloodmeals using mitochondrial cytochromes genes. *PloS one* **6**, e17284.

Ngo, H., Tschudi, C., Gull, K. and Ullu, E. (1998). Double-stranded RNA induces mRNA degradation in Trypanosoma brucei. Proceedings of the National Academy of Sciences of the United States of America **95**, 14687-14692.

Opperdoes, F. R., Butenko, A., Flegontov, P., Yurchenko, V. and Lukeš, J. (2016). Comparative Metabolism of Free-living *Bodo saltans* and Parasitic Trypanosomatids. *The J Eukaryot Microbiol* **63**, 657-678.

Pacheco, R. S., Marzochi, M. C., Pires, M. Q., Brito, C. M., Madeira Mde, F. and Barbosa-Santos, E. G. (1998). Parasite genotypically related to a monoxenous trypanosomatid of dog's flea causing opportunistic infection in an HIV positive patient. *Mem I Oswaldo Cruz* **93**, 531-537.

Patrick, K. L., Luz, P. M., Ruan, J. P., Shi, H., Ullu, E. and Tschudi, C. (2008). Genomic rearrangements and transcriptional analysis of the spliced leader-associated retrotransposon in RNA interference-deficient *Trypanosoma brucei*. *Mol Microbiol* **67**, 435-447.

Patrick, K. L., Shi, H., Kolev, N. G., Ersfeld, K., Tschudi, C. and Ullu, E. (2009). Distinct and overlapping roles for two Dicer-like proteins in the RNA interference pathways of the ancient eukaryote *Trypanosoma brucei*. *P Natl Acad Sci USA* **106**, 17933-17938.

Pays, E. and Vanhollebeke, B. (2008). Mutual self-defence: the trypanolytic factor story. *Microbes Infect* **10**, 985-989.

Peacock, C. S., Seeger, K., Harris, D., Murphy, L., Ruiz, J. C., Quail, M. A., Peters, N., Adlem, E., Tivey, A., Aslett, M., Kerhornou, A., Ivens, A., Fraser, A., Rajandream, M. A., Carver, T., Norbertczak, H., Chillingworth, T., Hance, Z., Jagels, K., Moule, S., Ormond, D., Rutter, S., Squares, R., Whitehead, S., Rabbinowitsch, E., Arrowsmith, C., White, B., Thurston, S., Bringaud, F., Baldauf, S. L., Faulconbridge, A., Jeffares, D., Depledge, D. P., Oyola, S. O., Hilley, J. D., Brito, L. O., Tosi, L. R., Barrell, B., Cruz, A. K., Mottram, J. C., Smith, D. F., Berriman, M. (2007). Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. *Nat Genet* **39**, 839-847.

Podlipaev, S. A. (1990). Podlipaev, S.A. (1990) Catalogue of world fauna of Trypanosomatidae (Protozoa). *Proc. Zool. Inst. USSR Acad. Sci.* **144**, 1-178.

Podlipaev, S. A., Sturm, N. R., Fiala, I., Fernandes, O., Westenberger, S. J., Dollet, M., Campbell, D. A. and Lukeš, J. (2004). Diversity of insect trypanosomatids assessed from the spliced leader RNA and 5S rRNA genes and intergenic regions. *The J Eukaryot Microbiol* **51**, 283-290.

Poinar, G., Jr. (2008). *Lutzomyia adiketis* sp. n. (Diptera: Phlebotomidae), a vector of *Paleoleishmania neotropicum* sp. n. (Kinetoplastida: Trypanosomatidae) in Dominican amber. *Parasit Vectors* **1**, 22.

Poinar, G., Jr. and Poinar, R. (2004). *Paleoleishmania proterus* n. gen., n. sp., (Trypanosomatidae: Kinetoplastida) from Cretaceous Burmese amber. *Protist* **155**, 305-310.

**Porcel, B. M., Denoeud, F., Opperdoes, F., Noel, B., Madoui, M. A., Hammarton, T. C., Field, M. C., Da Silva, C., Couloux, A., Poulain, J., Katinka, M., Jabbari, K., Aury, J. M., Campbell, D. A.,

Cintron, R., Dickens, N. J., Docampo, R., Sturm, N. R., Koumandou, V. L., Fabre, S., Flegontov, P., Lukes, J., Michaeli, S., Mottram, J. C., Szoor, B., Zilberstein, D., Bringaud, F., Wincker, P., Dollet, M. (2014). The streamlined genome of *Phytomonas* spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. *PLoS Genet* **10**, e1004007.

Pothirat, T., Tantiworawit, A., Chaiwarith, R., Jariyapan, N., Wannasan, A., Siriyasatien, P., Supparatpinyo, K., Bates, M. D., Kwakye-Nuako, G. and Bates, P. A. (2014). First isolation of *Leishmania* from Northern Thailand: case report, identification as *Leishmania martiniquensis* and phylogenetic position within the *Leishmania enriettii* complex. *PLOS Neglect Trop D* **8**, e3339.

Preusser, C., Rossbach, O., Hung, L. H., Li, D. and Bindereif, A. (2014). Genome-wide RNA-binding analysis of the trypanosome U1 snRNP proteins U1C and U1-70K reveals cis/trans-spliceosomal network. *Nucleic Acids Res* **42**, 6603-6615.

Rassi, A., Jr., Rassi, A. and Marin-Neto, J. A. (2010). Chagas disease. *Lancet* **375**, 1388-1402.

Real, F., Vidal, R. O., Carazzolle, M. F., Mondego, J. M., Costa, G. G., Herai, R. H., Wurtele, M., de Carvalho, L. M., Carmona e Ferreira, R., Mortara, R. A., Barbieri, C. L., Mieczkowski, P., da Silveira, J. F., Briones, M. R., Pereira, G. A., Bahia, D. (2013). The genome sequence of *Leishmania* (Leishmania) *amazonensis*: functional annotation and extended analysis of gene models. *DNA Res* **20**, 567-581.

Roberts, L., Janovy, J. and Nadler, S. (2013). *Foundations of Parasitology* (9th edition edn). New York, USA: McGraw-Hill.

Robinson, K. A. and Beverley, S. M. (2003). Improvements in transfection efficiency and tests of RNA interference (RNAi) approaches in the protozoan parasite *Leishmania*. *Mol Biochem Parasit* **128**, 217-228.

Runckel, C., DeRisi, J. and Flenniken, M. L. (2014). A draft genome of the honey bee trypanosomatid parasite *Crithidia mellificae*. *PloS One* **9**, e95057.

Ryan, K. J. and Ray, C. G. (2003). *Sherris Medical Microbiology*. New York, USA: McGraw Hill Professional.

Selvapandiyan, A., Ahuja, K., Puri, N. and Krishnan, A. (2015). Implications of co-infection of *Leptomonas* in visceral leishmaniasis in India. *Parasitology* **142**, 1657-1662.

Shaw, J. J. (1964). A Possible Vector of *Endotrypanum schaudinni* of the Sloth *Choloepus Hoffmanni*, in Panama. *Nature* **201**, 417-418.

Schwarz, R. S., Bauchan, G. R., Murphy, C. A., Ravoet, J., de Graaf, D. C. and Evans, J. D. (2015). Characterization of Two Species of Trypanosomatidae from the Honey Bee *Apis mellifera*: *Crithidia mellificae* Langridge and McGhee, and *Lotmaria passim* n. gen., n. sp. *The J Eukaryot Microbiol* **62**, 567-583.

Simao, F. A., Waterhouse, R. M., Ioannidis, P., Kriventseva, E. V. and Zdobnov, E. M. (2015). BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**, 3210-3212.

Simpson, A. G., Stevens, J. R. and Lukeš, J. (2006). The evolution and diversity of kinetoplastid flagellates. *Trends Parasitol* **22**, 168-174.

Singh, N., Chikara, S. and Sundar, S. (2013). SOLiD sequencing of genomes of clinical isolates of *Leishmania donovani* from India confirm leptomonas co-infection and raise some key questions. *PloS One* **8**, e55738.

Sistrom, M., Evans, B., Bjornson, R., Gibson, W., Balmer, O., Maser, P., Aksoy, S. and Caccone, A. (2014). Comparative genomics reveals multiple genetic backgrounds of human pathogenicity in the *Trypanosoma brucei* complex. *Genome Biol Evol* **6**, 2811-2819.

Srivastava, P., Prajapati, V. K., Vanaerschot, M., Van der Auwera, G., Dujardin, J. C. and Sundar, S. (2010). Detection of Leptomonas sp. parasites in clinical isolates of Kala-azar patients from India. *Infect Genet Evol* **10**, 1145-1150.

Stevens, J. R. and Gibson, W. (1999). The molecular evolution of trypanosomes. *Parasitol Today* **15**, 432-437.

Stevens, J. R., Noyes, H. A., Dover, G. A. and Gibson, W. C. (1999). The ancient and divergent origins of the human pathogenic trypanosomes, *Trypanosoma brucei* and *T. cruzi*. *Parasitology* **118 (Pt 1)**, 107-116.

Stork, N. E., McBroom, J., Gely, C. and Hamilton, A. J. (2015). New approaches narrow global species estimates for beetles, insects, and terrestrial arthropods. *P Natl Acad Sci USA* **112**, 7519-7523.

Svobodová, M., Zidková, L., Čepička, I., Oborník, M., Lukeš, J. and Votýpka, J. (2007). *Sergeia podlipaevi* gen. nov., sp. nov. (Trypanosomatidae, Kinetoplastida), a parasite of biting midges (Ceratopogonidae, Diptera). *Int J Syst Evol Micr* **57**, 423-432.

Tanifuji, G., Kim, E., Onodera, N. T., Gibeault, R., Dlutek, M., Cawthorn, R. J., Fiala, I., Lukeš, J., Greenwood, S. J. and Archibald, J. M. (2011). Genomic characterization of *Neoparamoeba pemaquidensis* (Amoebozoa) and its kinetoplastid endosymbiont. *Eukaryot Cell* **10**, 1143-1146.

Teixeira, M. M., Borghesan, T. C., Ferreira, R. C., Santos, M. A., Takata, C. S., Campaner, M., Nunes, V. L., Milder, R. V., de Souza, W. and Camargo, E. P. (2011). 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* **162**, 503-524.

Týč, J., Votýpka, J., Klepetková, H., Šuláková, H., Jirků, M., Lukeš, J. (2013). Growing diversity of trypanosomatid parasites of flies (Diptera: Brachycera): frequent cosmopolitism and moderate host specificity. Mol. Phyl. Evol. **69**, 255-264.

Verner, Z., Basu, S., Benz, C., Dixit, S., Dobakova, E., Faktorova, D., Hashimi, H., Horakova, E., Huang, Z., Paris, Z., Pena-Diaz, P., Ridlon, L., Tyc, J., Wildridge, D., Zikova, A., Lukes, J. (2015). Malleable mitochondrion of Trypanosoma brucei. Int Rev Cell Mol Biol **315**, 73-151.

Vickerman, K. (1962). The mechanism of cyclical development in trypanosomes of the *Trypanosoma brucei* sub-group: an hypothesis based on ultrastructural observations. *Trans R Soc Trop Med Hyg* **56**, 487-495.

Vickerman, K. (1985). Developmental cycles and biology of pathogenic trypanosomes. *Br Med Bull* **41**, 105-114.

Vickerman, K. (2000). Order Kinetoplastea. In *Illustrated Guide to the Protozoa* (ed. J. J. Lee, G. F. Leedale & P. C. Bradbury), pp. 1159-1184. Lawrence, Kansas, USA: American Society of Protozoologists.

Volf, P. and Horák, P. (2007). *Paraziti a jejich biologie*. Praha: Triton.

Votýpka, J., d'Avila-Levy, C. M., Grellier, P., Maslov, D. A., Lukeš, J. and Yurchenko, V. (2015). New Approaches to Systematics of Trypanosomatidae: Criteria for Taxonomic (Re)description. *Trends Parasitol* **31**, 460-469.

Votýpka, J., Klepetková, H., Jirků, M., Kment, P. and Lukeš, J. (2012a). Phylogenetic relationships of trypanosomatids parasitising true bugs (Insecta: Heteroptera) in sub-Saharan Africa. *Int J Parasitol* **42**, 489-500.

Votýpka, J., Klepetková, H., Yurchenko, V. Y., Horak, A., Lukeš, J. and Maslov, D. A. (2012b). Cosmopolitan distribution of a trypanosomatid *Leptomonas pyrrhocoris*. *Protist* **163**, 616-631.

Votýpka, J., Kostygov, A. Y., Kraeva, N., Grybchuk-Ieremenko, A., Tesarova, M., Grybchuk, D., Lukeš, J. and Yurchenko, V. (2014). *Kentomonas* gen. n., a new genus of endosymbiont-containing trypanosomatids of Strigomonadinae subfam. n. *Protist* **165**, 825-838.

Votýpka, J., Maslov, D. A., Yurchenko, V., Jirků, M., Kment, P., Lun, Z. R. and Lukeš, J. (2010). Probing into the diversity of trypanosomatid flagellates parasitizing insect hosts in South-West China reveals both endemism and global dispersal. *Mol Phylogenet Evol* **54**, 243-253.

Votýpka, J., Rádrová, J., Skalický, T., Jirků, M., Jirsová, D., Mihalca, A. D., D'Amico, G., Petrželková, K. J., Modrý, D. and Lukeš, J. (2015). A tsetse and tabanid fly survey of African great apes habitats reveals the presence of a novel trypanosome lineage but the absence of *Trypanosoma brucei*. *Int J Parasitol* **45**, 741-748.

Votýpka, J., Suková, E., Kraeva, N., Ishemgulova, A., Duzi, I., Lukeš, J. and Yurchenko, V. (2013). Diversity of trypanosomatids (Kinetoplastea: Trypanosomatidae) parasitizing fleas (Insecta: Siphonaptera) and description of a new genus *Blechomonas* gen. n. *Protist* **164**, 763-781.

Wallace, F. G. (1977). *Leptomonas seymouri* sp. n. from the cotton stainer *Dysdercus suturellus*. *J Protozool* **24**, 483-484.

Wallace, F. G., Camargo, E. P., McGhee, R. B. and Roitman, I. (1983). Guidelines for the description of new species of lower trypanosomatids. *J. Eukaryot. Microbiol.* **30**, 308–313.

Westenberger, S. J., Sturm, N. R., Yanega, D., Podlipaev, S. A., Zeledon, R., Campbell, D. A. and Maslov, D. A. (2004). Trypanosomatid biodiversity in Costa Rica: genotyping of parasites from Heteroptera using the spliced leader RNA gene. *Parasitology* **129**, 537-547.

Wheeler, R. J., Gluenz, E. and Gull, K. (2015). Basal body multipotency and axonemal remodelling are two pathways to a 9+0 flagellum. *Nature Commun* **6**, 8964.

Wheeler, R. J., Scheumann, N., Wickstead, B., Gull, K. and Vaughan, S. (2013). Cytokinesis in *Trypanosoma brucei* differs between bloodstream and tsetse trypomastigote forms: implications for microtubule-based morphogenesis and mutant analysis. *Mol Microbiol* **90**, 1339-1355.

Wheeler, R. J., Sunter, J. D. and Gull, K. (2016). Flagellar pocket restructuring through the *Leishmania* life cycle involves a discrete flagellum attachment zone. *J Cell Sci* **129**, 854-867.

Woo, P. T. (2003). *Cryptobia* (*Trypanoplasma*) *salmositica* and salmonid cryptobiosis. *J Fish Dis* **26**, 627-646.

Worthey, E. A., Martinez-Calvillo, S., Schnaufer, A., Aggarwal, G., Cawthra, J., Fazelinia, G., Fong, C., Fu, G., Hassebrock, M., Hixson, G., Ivens, A. C., Kiser, P., Marsolini, F., Rickel, E., Salavati, R., Sisk, E., Sunkin, S. M., Stuart, K. D., Myler, P. J. (2003). *Leishmania major* chromosome 3 contains two long convergent polycistronic gene clusters separated by a tRNA gene. *Nucleic Acids Res* **31**, 4201-4210.

Young, N. D., Dykova, I., Crosbie, P. B., Wolf, M., Morrison, R. N., Bridle, A. R. and Nowak, B. F. (2014). Support for the coevolution of *Neoparamoeba* and their endosymbionts, *Perkinsela* amoebae-like organisms. *Eur J Protistol* **50**, 509-523.

Yurchenko, V. Y., Lukeš, J., Jirků, M. and Maslov, D. A. (2009). Selective recovery of the cultivation-prone components from mixed trypanosomatid infections: a case of several novel species isolated from Neotropical Heteroptera. *Int J Syst Evol Micr* **59**, 893-909.

Yurchenko, V. Y., Lukeš, J., Tesarova, M., Jirků, M. and Maslov, D. A. (2008). Morphological discordance of the new trypanosomatid species phylogenetically associated with the genus C*rithidia*. *Protist* **159**, 99-114.

Záhonová, K., Kostygov, A. Y., Ševčíková, T., Yurchenko, V. and Eliáš, M. (2016). An Unprecedented non-canonical nuclear genetic code with all three termination codons reassigned as sense codons. *Curr Biol* **26**, 2364-2369.

5 Original publications

5.1 PAPER I

Flegontov P, Votýpka J, **Skalický T**, Logacheva MD, Penin AA, Tanifuji G, Onodera NT, Kondrashov AS, Volf P, Archibald JM, Lukeš J (2013) *Paratrypanosoma* is a novel early-branching trypanosomatid. **Current Biology** 23(18):1787-93 (IF = 9.733).

Abstract

The kinetoplastids are a widespread and important group of single-celled eukaryotes, many of which are devastating parasites of animals, including humans. We have discovered a new insect trypanosomatid in the gut of *Culex pipiens* mosquitoes. Glyceraldehyde-3-phosphate dehydrogenase- and SSU rRNA-based phylogenetic analyses show this parasite to constitute a distinct branch between the free-living *Bodo saltans* and the obligatory parasitic clades represented by the genus *Trypanosoma* and other trypanosomatids. From draft genome sequence data, we identified 114 protein genes shared among the new flagellate, 15 trypanosomatid species, *B. saltans*, and the heterolobosean *Naegleria gruberi*, as well as 129 protein genes shared with the basal kinetoplastid *Perkinsela* sp. Individual protein phylogenies together with analyses of concatenated alignments show that the new species, here named *Paratrypanosoma confusum* n. gen., n. sp., branches with very high support at the base of the family Trypanosomatidae. *P. confusum* thus represents a long-sought-after missing link between the ancestral free-living bodonids and the derived parasitic trypanosomatids. Further analysis of the *P. confusum* genome should provide insight into the emergence of parasitism in the medically important trypanosomatids.

Current Biology 23, 1787–1793, September 23, 2013 ©2013 Elsevier Ltd All rights reserved http://dx.doi.org/10.1016/j.cub.2013.07.045

Report

Paratrypanosoma Is a Novel Early-Branching Trypanosomatid

Pavel Flegontov,[1,9] Jan Votýpka,[1,2,9] Tomáš Skalický,[1,3,9] Maria D. Logacheva,[4,5] Aleksey A. Penin,[4,6] Goro Tanifuji,[7] Naoko T. Onodera,[7] Alexey S. Kondrashov,[4,8] Petr Volf,[2] John M. Archibald,[7] and Julius Lukeš[1,3,*]
[1]Institute of Parasitology, Biology Centre, 37005 České Budějovice (Budweis), Czech Republic
[2]Department of Parasitology, Faculty of Science, Charles University, 12844 Prague, Czech Republic
[3]Faculty of Science, University of South Bohemia, 37005 České Budějovice (Budweis), Czech Republic
[4]Faculty of Bioengineering and Bioinformatics, Lomonosov Moscow State University, Moscow 119991, Russia
[5]A.N. Belozersky Institute of Physico-Chemical Biology, Lomonosov Moscow State University, Moscow 119991, Russia
[6]Kharkevich Institute for Information Transmission Problems, Russian Academy of Sciences, Moscow 127994, Russia
[7]Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, NS B3H 4R2, Canada
[8]Life Sciences Institute and Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, MI 48109, USA

Summary

The kinetoplastids are a widespread and important group of single-celled eukaryotes, many of which are devastating parasites of animals, including humans [1–3]. We have discovered a new insect trypanosomatid in the gut of *Culex pipiens* mosquitoes. Glyceraldehyde-3-phosphate dehydrogenase- and SSU rRNA-based phylogenetic analyses show this parasite to constitute a distinct branch between the free-living *Bodo saltans* and the obligatory parasitic clades represented by the genus *Trypanosoma* and other trypanosomatids. From draft genome sequence data, we identified 114 protein genes shared among the new flagellate, 15 trypanosomatid species, *B. saltans*, and the heterolobosean *Naegleria gruberi*, as well as 129 protein genes shared with the basal kinetoplastid *Perkinsela* sp. Individual protein phylogenies together with analyses of concatenated alignments show that the new species, here named *Paratrypanosoma confusum* n. gen., n. sp., branches with very high support at the base of the family Trypanosomatidae. *P. confusum* thus represents a long-sought-after missing link between the ancestral free-living bodonids and the derived parasitic trypanosomatids. Further analysis of the *P. confusum* genome should provide insight into the emergence of parasitism in the medically important trypanosomatids.

Results

Isolation and Morphological and Ultrastructural Characterization

Out of 206 female mosquitoes (*Culex pipiens*) captured in Prague in June 2000, 25 were found to contain flagellates in their intestine. In most cases, these infections were confined to the midgut and stomodeal valve, characteristic for avian trypanosomes [4]. However, in the midgut and hindgut of several mosquitoes, different slowly moving flagellates, which were in one case successfully established as an axenic culture in SNB medium, were observed and were upon further study named *Paratrypanosoma confusum* n. gen., n. sp. (see below and "Taxonomic Summary").

The *P. confusum* culture is dominated by elongated promastigote-shaped cells, defined by the mutual position of the nucleus and kinetoplast DNA (kDNA) (Figure 1A). Occasionally, ovoid stages with morphology reminiscent of choanomastigotes occur (Figure 1B). Transmission electron microscopy (Figures 1C–1I) shows that the plasmalemma is underlain by a complete corset of subpellicular microtubules (Figures 1C and 1G). All kDNA is packed in a single dense disk, with minicircles stretched taut, located in the canonical position at the base of the single flagellum (Figures 1F and 1I). Considering the known correlation between the thickness of the disk and size of kDNA minicircles, their size in *P. confusum* is estimated to be 0.8 kb. The nucleus is usually located in or close to the center of the cell (Figures 1A, 1H, and 1I), and, as in other trypanosomatids, the kDNA division predates nuclear division (Figure 1H). A single long flagellum is supported by a prominent paraflagellar rod (Figure 1E), which is absent in the flagellar pocket (Figure 1D). Numerous vesicles reminiscent of acidocalcisomes and lipid bodies (Figures 1A–1C,1G–1I) are present throughout the cell.

SSU rRNA and gGAPDH Phylogeny

Nuclear small subunit (SSU) ribosomal RNA (rRNA) sequences were amplified from *P. confusum* DNA isolated from the axenic culture. In addition, identical or very similar partial SSU rRNA sequences have been amplified by nested PCR in monospecies pools of female *C. pipiens* and *C. modestus* mosquitoes trapped in southern Bohemia (Řežabinec) and southern Moravia (Mikulov), Czech Republic (data not shown). Short SSU rRNA fragments of 345 bp (accession numbers DQ813272–DQ813295) matching the *P. confusum* sequence with 95%–100% identity (data not shown) were previously amplified from *C. pipiens* and *C. tarsalis* mosquitoes collected in Colorado [5]. The *P. confusum* sequence and 219 nonidentical bodonid and trypanosomatid SSU rRNA sequences were aligned using SINA aligner [6] and were manually edited, resulting in 1,325 aligned characters (Figure S1A available online). The resulting maximum likelihood (ML) tree shows that *P. confusum* is clearly distinct from all known trypanosomatid clades: *Trypanosoma* [7], *Blastocrithidia-Leptomonas jaculum* [1, 8], *Herpetomonas* [9], *Phytomonas* [1], *Angomonas-Strigomonas* [10], *Sergeia*, *Leptomonas collosoma* [1, 8], and subfamily Leishmaniinae [11]. *P. confusum* is the most basal trypanosomatid branch with 98% bootstrap support (Figures 2 and S2A).

Next, the glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) gene was amplified and sequenced. An amino acid sequence alignment of 294 characters was constructed using 143 bodonid and trypanosomatid sequences (Figure S1B); phylogenetic model selection using Modelgenerator favored LG+Γ as the best model for this alignment. The

[9]These authors contributed equally to this work
*Correspondence: jula@paru.cas.cz

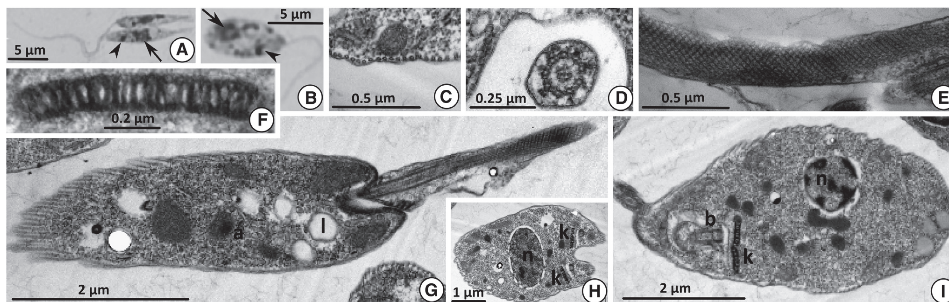


Figure 1. Morphology and Ultrastructure of *Paratrypanosoma confusum* n. sp.

(A and B) Light microscopy images. Giemsa staining of the predominant slender promastigotes (A) and infrequent oval-shaped promastigotes (B) reveals the position of the nucleus (arrow) and kinetoplast DNA (kDNA; arrowhead).
(C–I) Transmission electron microscopy images. k, kDNA; n, nucleus; b, basal body; l, lipid granule; a, acidocalcisome.
(C) The plasmalemma is supported by a corset of subpellicular microtubules.
(D) Section of the single flagellum within the flagellar pocket lacks the paraflagellar rod.
(E) An intricate meshwork of the prominent paraflagellar rod.
(F) Thin and wide kDNA disk composed of minicircles stretched taut.
(G) Longitudinally sectioned promastigote revealing full corset of subpellicular microtubules, external paraflagellar rod, and numerous lipid granules and putative acidocalcisomes.
(H) Division of the kDNA precedes nuclear division.
(I) Longitudinal section of an oval-shaped promastigote, revealing the relative position of the nucleus and the kDNA disk.

topology of the gGAPDH-based tree, which is, when compared with the SSU rRNA data set, underrepresented for the monoxenous lineages, further supports separation of *P. confusum* from the other trypanosomatid clades (bootstrap support 88%) (Figure S2B).

Phylogenomics

In order to further explore the possibility of *P. confusum* belonging to a novel trypanosomatid lineage, we used next-generation sequencing to produce a draft genome sequence. A paired-end Illumina library (insert size 330 ± 50 bp) was prepared from *P. confusum* total DNA and sequenced on a HiSeq 2000 instrument. Assembly of 40.1 million quality-filtered 101 bp reads gave scaffold N50 value of 11,534 bp. This assembly was used as a database from which to harvest protein genes for subsequent phylogenomic analyses.

Translated open reading frames (ORFs) >100 amino acids in length, from AUG to stop codons, were extracted from the *P. confusum* assembly. Simple ORF finding was used instead of more-sophisticated annotation methods because kinetoplastid genomes are essentially devoid of introns [2, 12]. Best reciprocal BLASTP hits at an E-value cutoff of 10^{-20} were found for *P. confusum* ORFs in annotated proteins or translated ORFs of 15 trypanosomatid species and the free-living bodonid *B. saltans* (Table S1, part A). Proteins from *Naegleria gruberi* (Excavata) (15,762 NCBI RefSeq entries) were used as outgroups. Proteins inferred from transcriptome data from the basal branching kinetoplastid *Perkinsela* sp. CCAP 1560/4, an endosymbiont of the amoebozoan *Neoparamoeba pemaquidensis* (G.T., P.F., N.T.O., J.L., and J.M.A., unpublished data), were included as outgroup sequences in some data sets.

Sequence clusters containing all 17 kinetoplastid species/strains and an outgroup(s) were selected for further analysis. Clusters including sequences more than two times longer or shorter than the average for a given cluster were excluded, as were clusters including sequences with BLASTP hit length 1.5 times longer or shorter than average for the cluster and/or with an average identity in the BLASTP hit region <40% between an outgroup and the other species. The following six data sets were generated: (1) 114 sequence clusters with *N. gruberi* as an outgroup, including alignments with gaps if occurring in less than half of the sequences; (2) the same alignments as in (1) without gaps; (3) 129 sequence clusters with *Perkinsela* sp. as an outgroup; (4) the same alignments as in (3) without gaps; (5) 42 sequence clusters with both *N. gruberi* and *Perkinsela* sp.; and (6) the same alignments as in (5) without gaps (Tables S2 and S3; alignments are available upon request). Phylogenetic model selection with Modelgenerator favored the LG+Γ+F model for all six concatenated alignments, and ML trees were constructed with 1,000 bootstrap replicates using this model or the GTR+Γ model (Figure 3; Table S2, part A). In all six data sets, *P. confusum* branched between *B. saltans* and the genus *Trypanosoma* with 100% bootstrap support. The branching order of the other trypanosomatid species matched the expected pattern [1], with all nodes having 100% bootstrap support. Bayesian Monte Carlo Markov (MCM) chain analysis of the concatenated data sets, conducted with the Poisson+Γ+CAT or GTR+Γ+CAT models, showed the same branching position for *P. confusum*, with posterior probabilities ranging from 0.95 (gapped data set with *N. gruberi* and *Perkinsela* sp., Poisson+Γ+CAT model) to 1 (all data sets with GTR+Γ+CAT model; both data sets with *Perkinsela* sp., Poisson+Γ+CAT model) (Figure 3; Table S2, part A). Convergence of chains was estimated by comparison of bipartition frequencies in individual chains, discarding the first 2,000 cycles and taking each tree; maximum difference in frequencies among chains ranged from 0.11 (data set with *N. gruberi*, no gaps, Poisson+Γ+CAT model) to 0 (all data sets with GTR+Γ+CAT model;

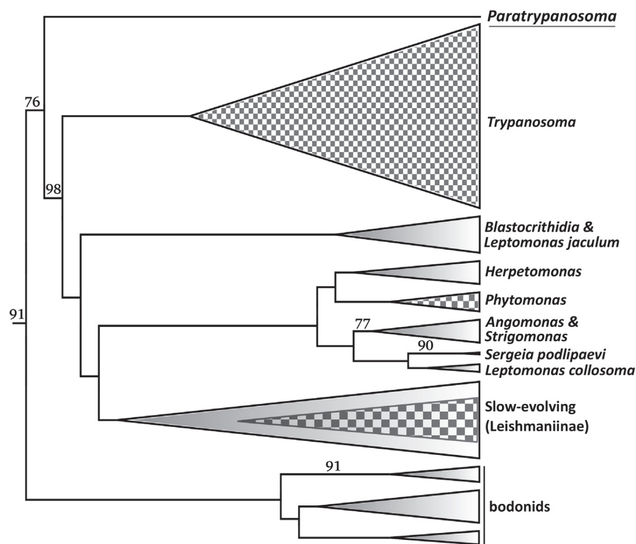


Figure 2. Maximum-Likelihood Phylogeny Based on the SSU rRNA Gene, Constructed under the GTR+Γ Model

Simplified representation; for the full tree, see Figure S2A. Clades containing dixenous species are highlighted with checkerboard shading. *P. confusum* is underlined. Bootstrap values >75% are displayed. See also Figures S1 and S2.

both data sets with *Perkinsela* sp., Poisson+Γ+CAT model). Groupings in conflict with the most probable topology corresponded to trees in which *P. confusum* branched specifically with *B. saltans* or with the outgroup (Table S2, part A). The GTR+Γ+CAT model performed better than the Poisson+Γ+CAT model according to cross-validation tests using data sets without gaps and resulted in perfect convergence in all data sets (Table S2, part A).

Removal of the BLASTP hit identity cutoff of 40% expanded the data set to 226 proteins when *N. gruberi* was used as an outgroup (Table S1, part B). Phylogenetic model selection with Modelgenerator favored the LG+Γ+F for the gapped and ungapped concatenated alignments, and ML trees for both alignments were constructed using this model (Table S2, part A). The results supported the position of *P. confusum* at the base of trypanosomatids in all bootstrap replicates. Bayesian analysis of this data set with the Poisson+Γ+CAT model was compromised by poor chain convergence, probably due to increased long-branch attraction (LBA) effects in a data set containing less conserved proteins (Table S2, part A). On the other hand, selection of more conserved proteins with a stricter BLASTP hit identity cutoff of 50% (data set with *N. gruberi*, 52 proteins; Table S1, part B) did not change the ML tree topology and support and decreased the frequency of conflicting bipartitions in the MCM chains under both the Poisson+Γ+CAT and the GTR+Γ+CAT models (Table S2, part A). The GTR+Γ+CAT had better fit than the Poisson+Γ+CAT model according to cross-validation tests with "*N. gruberi* 52" data sets (Table S2, part A).

Posterior predictive analyses of mutational saturation under the Poisson+Γ+CAT model showed that the numbers of substitutions and homoplasies were not underestimated. As expected, their numbers per site, six and three, respectively, were the lowest for the most "conserved" data set (with *N. gruberi*; 52 proteins, no gaps) (Table S2, part B). Mutational saturation in this data set is on the same level as in the metazoan alignment used for inferring the placement of nematodes and platyhelminths under the Poisson+Γ+CAT model: 7.75 and 4.3 substitutions and homoplasies per site, respectively [13]. Under these conditions, the model predicted mutational saturation correctly and was therefore deemed to be resistant to LBA, as opposed to the WAG+Γ+F model [13].

Single-protein ML trees were constructed for the 52, 114, and 129 protein data sets with either *N. gruberi* or *Perkinsela* sp. as an outgroup. The topology in which *P. confusum* branches between *B. saltans* and the other trypanosomatids was the most frequently observed, consistent with the results of concatenated analyses. However, *P. confusum* forming a monophyletic group with *B. saltans* (located at different positions on the tree) was the second most frequent topology (Table S3). Other topologies in order of decreasing frequency were (1) *P. confusum* as the sister branch of the *Trypanosoma* clade only, (2) *P. confusum* branching before *B. saltans*, and (3) *P. confusum* as the sister branch of the subfamily Leishmaniinae and *Phytomonas* clade only (Table S3). The branching of *P. confusum* with *B. saltans* or deeper in the tree than *B. saltans* is most probably the result of LBA, and indeed such topologies are more frequent in single-protein trees derived from the mutationally saturated (Table S2, part B) data set of 226 proteins (Table S3). In contrast, the branching of *P. confusum* with the genus *Trypanosoma* is less obviously an artifact. While avian trypanosomes were also isolated from *Culex* mosquitoes [4, 5], *P. confusum* clearly lacks the trypomastigote morphology synapomorphic for the genus *Trypanosoma* [1]. We used topology tests to examine the possibility of a specific relationship between *P. confusum* and trypanosomes.

In the "*N. gruberi* 114," "*N. gruberi* 52," and "*Perkinsela* sp. 129" gapped and ungapped data sets, topologies within eight important clades on the tree were fixed: the outgroup, *B. saltans*, *P. confusum*, *Trypanosoma* spp., *Phytomonas serpens*, *Crithidia fasciculata* + *Leptomonas pyrrhocoris*, and finally *Leishmania* spp. + *Endotrypanum monterogeii*. All possible 10,395 topologies of the seven clades rooted with the outgroup were constructed for each data set, and per-site log likelihoods were calculated for all topologies under the LG+Γ+F or GTR+Γ phylogenetic models. In all cases, the approximately unbiased (AU) test did not support the grouping of *P. confusum* with *Trypanosoma* or with Leishmaniinae and *Phytomonas* at a p value cutoff of 10^{-4} (Table S4, part A).

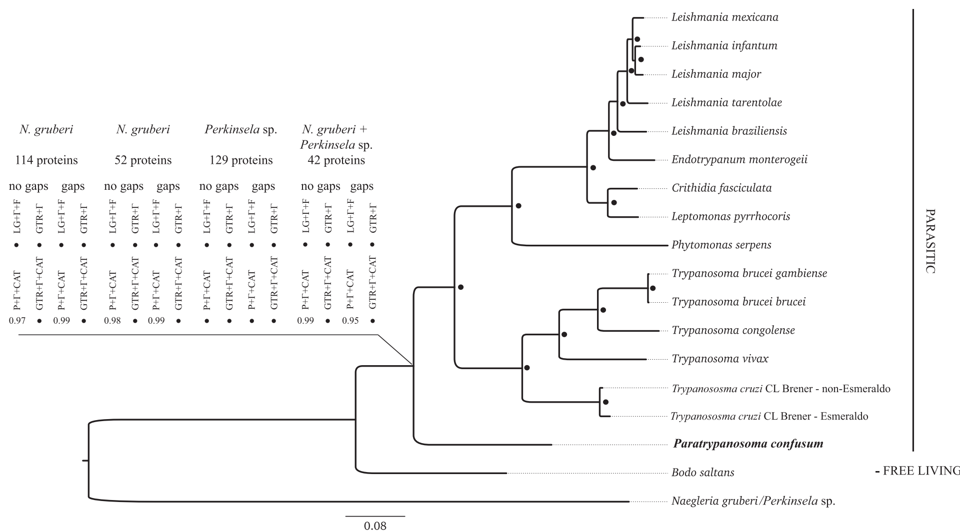


Figure 3. Maximum-Likelihood Phylogenetic Tree Based on Concatenated Protein Alignments of 18 Species

Nodes having 100% bootstrap support or posterior probabilities of 1.0 in all data sets and under all phylogenetic models tested are marked by black circles. Support for the position of *P. confusum* (in bold) is shown in a separate table. The scale bar indicates the inferred number of amino acid substitutions per site. See also Tables S1 and S2.

For most data sets, none of the 10,394 alternative topologies were supported by the AU test at a p value cutoff of 0.05. Exceptions included only topologies strongly conflicting previous data, such as the branching of *P. confusum* basal to *B. saltans* and the branching of *L. major* basal to *C. fasciculata* and *P. serpens* (Table S4, part B).

Taxonomic Summary

The taxonomic summary of *Paratrypanosoma confusum* n. gen., n. sp., is as follows:

- Class Kinetoplastea Honigberg, 1963 emend. Vickerman, 1976
- Subclass Metakinetoplastina Vickerman, 2004
- Order Trypanosomatida Kent, 1880 stat. nov. Hollande, 1952
- Family Trypanosomatidae Doflein, 1951

Paratrypanosomatinae n. subfam. Votýpka and Lukeš 2013

The newly described subfamily belongs to the obligatory parasitic uniflagellate family Trypanosomatidae, with kDNA arranged in a single compact disk at the base of the flagellum. Diagnosis is phylogenetically defined by branching at the base of all trypanosomatids, according to SSU rRNA and multiple protein-coding genes. The type genus is *Paratrypanosoma* n. gen. Votýpka and Lukeš 2013.

Paratrypanosoma confusum n. sp. Votýpka and Lukeš 2013

The dominant morphotype observed in the axenic culture is an elongated promastigote, 9.8 ± 2.1 (7.2–16.2) μm long and 2.0 ± 0.3 (1.5–2.9) μm wide, with 12.2 ± 4.8 (7.0–39.9) μm long flagellum (n = 50). The nucleus and the kDNA are situated in the anterior end of the cell. The distance between the anterior end and the kDNA and the nucleus is 2.3 ± 0.3 (1.5–2.9) μm and 4.6 ± 0.7 (3.3–6.2) μm, respectively (n = 50). The thickness of the kinetoplast is 116.4 ± 11.7 (94.4–155.5) nm (n = 50). Short oval promastigotes of varying sizes were rare.

Type host and locality: intestine of mosquito female *C. pipiens* (Diptera: Nematocera: Culicidae) captured on June 28, 2000 in the vicinity of Prague-Prosek (50°6′45.81″N, 14°29′14.53″E).

Additional hosts and localities: female mosquitoes *Culex pipiens* and *C. modestus* in Řežabinec (49°15′09″N, 14°05′32″E) and Mikulov (48°46′30″N, 16°43′30″E), Czech Republic.

Type material: the designated hapantotype is cryopreserved as an axenic culture of *P. confusum* (isolate CUL13) deposited in the slide collection at Charles University, Prague.

Etymology: the species name was given to reflect the misleading morphology.

Gene sequences: The GenBank accession numbers are KC534633–KC534828.

Discussion

Using morphology and phylogenomics, we have described a new kinetoplastid, *Paratrypanosoma confusum*, which constitutes the most basal trypanosomatid lineage branching between the free-living *B. saltans* and the parasitic *Trypanosoma* spp. and other trypanosomatids. Kinetoplastid

flagellates (Kinetoplastea, Euglenozoa) are ubiquitous single-celled eukaryotes best known as pathogens of humans and other animals, responsible for African sleeping sickness, Chagas disease, leishmaniases, and other diseases. They are traditionally split into the bodonids, which are comprised of biflagellate free-living, commensalic, or parasitic members, and the obligatory parasitic trypanosomatids, which are equipped with a single flagellum [2, 14]. Bodonids and trypanosomatids also share some unusual molecular features, such as packaging of kDNA, RNA editing, polycistronic transcription, highly modified base J, and massive *trans*-splicing [12, 14, 15]. Extensive phylogenetic analyses of about a dozen bodonid and more than a hundred trypanosomatid species have shown that the latter group is monophyletic, whereas bodonids are clearly paraphyletic [14, 15].

The origin of the extremely successful trypanosomatid life style, which combines a vertebrate (usually warm-blooded) host with an invertebrate (usually insect) vector, has been debated for more than a century [16, 17]. The insect-early scenario is now generally favored [1], since phylogenies constructed from multiple nuclear-encoded proteins suggest that the dixenous (two-host) genera *Leishmania* and *Phytomonas* are nested within clades that otherwise consist of monoxenous (single-host) insect trypanosomatids [1, 8, 11, 18]. Recent molecular surveys uncovered a major hidden diversity of insect trypanosomatids, greatly exceeding that of the dixenous genera [8–11, 18, 19]; globally, more than 10% of all dipterans, fleas, and hemipterans may be infected [1].

Hence, the most likely scenario for the evolution of dixenous parasitism postulates that an ancestor of *Leishmania* parasitizing a blood-sucking insect was injected into a vertebrate host during blood feeding and established itself in that niche. This course of events is supported by the discovery of an amber-trapped phlebotomine sand fly that was massively infected by flagellates virtually indistinguishable from the extant *Leishmania*; the insect's intestinal tract also contained nucleated red blood cells, likely originating from a "dinosaur" [20]. The protist was dated to ~220 million years ago, indicating that the establishment of the dixenous life cycle may be a fairly ancient event [15, 20]. Phylogenetic position of *Phytomonas* favors a similar scenario, in which flagellates established themselves in plants only after being transmitted to them by infected sap-sucking insects [1].

The third group known to have adopted a dixenous life style is the emblematic genus *Trypanosoma*, which thrives in a wide variety of hosts, ranging from deep-sea fish and desert reptiles to birds and mammals, including humans [2, 3, 7]. Trypanosomes have been extensively studied since Bruce's discovery of sleeping sickness [21], and their diversity is fairly well known [3, 7]. With the advent of molecular techniques, it was shown that the genus *Trypanosoma* constitutes the most basal trypanosomatid branch, the monophyly of which withstood phylogenetic scrutiny, yet sometimes its early-branching position could not be resolved with confidence [1–3, 7, 22]. However, since the time of Léger and Minchin [16, 17], the search for monoxenous ancestors of *Trypanosoma* has been ongoing, which would illuminate the evolution of the *Trypanosoma* life cycle and emergence of its extremely successful parasitic strategy.

P. confusum is the first flagellate to fit this bill. Due to its origin from female mosquitoes, its monoxenous status may be questioned, but since it was repeatedly encountered in mosquitoes in the Czech Republic (this work) and US [5], yet was never found in much-better-studied vertebrates, we consider its transmission to the latter hosts to be highly unlikely. Our extensive phylogenetic analyses argue against a specific sister relationship between *P. confusum* and *Trypanosoma* spp. and other trypanosomatids with genome sequences available. The free-living biflagellate *B. saltans*, which is radically different from parasitic species in morphology and biology, represents the closest known outgroup to *P. confusum* [12, 14, 15]. Thus, detailed genome sequence analysis of *P. confusum* should provide crucial information about the switch to a parasitic life style followed by the uniquely successful expansion of trypanosomatid flagellates.

P. confusum promastigote morphology, which occurs in various trypanosomatid clades, may be considered ancestral for the group as a whole, while the trypomastigote and epimastigote morphologies represent a synapomorphy of the genus *Trypanosoma*. Significantly, the main ultrastructural characters shared by all trypanosomatids are already present in *P. confusum*.

Since all SSU rRNA sequences obtained for *P. confusum* on two continents (this work; [5]) are virtually identical, strongly indicating their monospecific character, *Paratrypanosoma* may be an example of a cosmopolitan yet species-poor clade. Most surveys targeting non-*Leishmania* and non-*Trypanosoma* flagellates involved only two insect orders, Heteroptera and Diptera, with very few reports available from other groups, such as Hymenoptera [23] and Siphonaptera (J.V. and J.L., unpublished data; [1]). To what extent this reflects the actual distribution of parasites in insects, or simply investigator bias, remains unclear. In any case, the finding of the most deeply diverging branch in a dipteran host suggests that association of trypanosomatids with this insect order may be an ancestral state, with its spread to other insects, plants and vertebrates occurring secondarily.

Experimental Procedures

Collection of Insects, Cultivation, Microscopy, and DNA Isolation

Mosquito females were collected by miniature Centers for Disease Control (CDC) traps when attacking sparrow-hawk (*Accipiter nisus*) nestlings in Prague [4] or by dry ice- or animal-baited CDC traps in several wetland regions in southern Bohemia and Moravia. Mosquitoes were dissected under a stereomicroscope, their alimentary tracts were examined using light microscopy, and axenic culture was established as described previously [4]. The cultures were kept at 23°C, and cells were processed for light and transmission electron microscopy as described elsewhere [11].

Total DNA was isolated from axenic *P. confusum* culture or from ethanol-preserved environmental samples of mosquito females grouped in monospecific pools using the High Pure PCR Template Preparation Kit (Roche) according to the manufacturer's manual. SSU rRNA and glycosomal GAPDH genes were amplified using primers S762 (5′-GACTTTTGCTTCC TCTA[A/T]TG-3′) and S763 (5′-CATATGCTTGTTTCAAGGAC-3′) and M200 (5′-ATGGCTCC[G/A/C][G/A/C]TCAA[G/A]GT[A/T]GG[A/C]AT-3′) and M201 (5′-TA[G/T]CCCCACTCGTT[G/A]TC[G/A]TACCA-3′), respectively. Upon gel purification with the Gel Extraction Kit (Roche), the amplicons were directly sequenced.

High-Throughput DNA Sequencing and Sequence Assembly

To construct the libraries for whole-genome sequencing, DNA was processed as described in the TruSeq DNA Sample Preparation Guide (Illumina). The library with length of 330 ± 50 bp (according to analysis on Agilent 2100 Bioanalyzer) was selected for sequencing. The library was quantified using fluorimetry with Qubit (Invitrogen) and real-time PCR and diluted up to final concentration of 8 pM. The diluted library was clustered on a paired-end flowcell (TruSeq PE Cluster Kit v3) using a cBot instrument and sequenced using a HiSeq2000 sequencer with the TruSeq SBS Kit v3-HS with a read length of 101 bp from each end. The following bases/reads were removed at the filtering stage using PRINSEQ: 3′ tails with Phred

quality values (QVs) <3, homopolymeric 3′ tails >10 nt, reads with length <75 nt, reads with average base QV <30, and reads with more than one undetermined base. Genome assembly was performed using Velvet 1.2.03 with a range of assembly parameters tested: odd k-mer values from 25 to 49; k-mer coverage cutoffs 3, 6, or 9; and expected k-mer coverage from 60 to 10,000. The best scaffold N50 value of 11,534 bp was obtained with a k-mer value of 27, coverage cutoff 6, and expected coverage 10,000.

Phylogenetic Analyses

The SSU rRNA alignment was constructed using the SINA aligner website [6] and edited manually; the ML tree for SSU rRNAs was constructed with RAxML 7.2.8 using the GTR+Γ model, with 1,000 bootstrap replicates. For the gGAPDH alignment, the ML tree was constructed using the LG+Γ model and 1,000 bootstrap replicates. Protein alignments were made using MUSCLE 3.8.31 with maximum number of iterations set to 128. Site selection was performed with Gblocks 0.91b [24] using the following settings: minimum number of sequences for a conserved position, ten; minimum number of sequences for a flanking position, 12; maximum number of contiguous nonconserved positions, eight; minimum length of a block, ten; and allowed gap positions, in half of sequences or none. Selection of appropriate model parameters was done with Modelgenerator [25] using four Γ rate categories. Single-protein ML trees were constructed using RAxML 7.2.8, the LG+Γ model and 100 bootstrap replicates in all cases. For multiprotein concatenated alignments, the ML trees were constructed using different phylogenetic models and 1,000 bootstrap replicates (identical random seeds were used in all analyses), and Bayesian phylogenetic trees were constructed using PhyloBayes 3.3f, with eight chains run for 10,000 cycles. Posterior predictive tests of substitutions and homoplasies were performed using PhyloBayes 3.3f. Cross-validation of the Poisson+Γ+CAT model versus the reference GTR+Γ+CAT model was performed using the PhyloBayes package as follows: ten replicates of each data set were split into a learning set (nine-tenths of the initial data set size) and a test set (one-tenth of the initial data set size); each model was run with each replicated learning set for 10,000 cycles under the topology estimated by the model itself on the full data set; and cross-validated log-likelihood scores of each test set were computed, taking each tree after discarding first 2,000 trees, and combined. For the purpose of topology testing, lists of all possible topologies were generated using Tresolve (Barrel-o-Monkeys package, http://rogerlab.biochemistryandmolecularbiology.dal.ca/Software/Software.htm), and per-site log likelihoods for the topologies were calculated using RAxML 7.2.8. Bootstrap replicates of per-site log likelihoods were made using CONSEL 0.1i with the "-b 10" option producing ten data sets of scales 0.5, 0.6, 0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, and 1.4, with 100,000 replicates in each. AU test p values and their SEs were calculated using CONSEL 0.1i [26].

Accession Numbers

The GenBank accession numbers KC534504–KC534632 for the transcriptomic data reported in this paper were obtained by sequencing *Perkinsela* sp. strain CCAP1560/4. The GenBank accession numbers KC534633–KC534828 for the genomic data reported in this paper were obtained by sequencing *Paratrypanosoma confusum*. The GenBank accession numbers KC543586–KC543700 for the genomic data reported in this paper were obtained by sequencing *Leptomonas pyrrhocoris* strain H10.

Supplemental Information

Supplemental Information includes two figures and four tables and can be found with this article online at http://dx.doi.org/10.1016/j.cub.2013.07.045.

Acknowledgments

The help of Eva Suková with photodocumentation, Aleš Horák with phylogenetic analyses, and Milena Svobodová and Jana Rádrová with isolation and PCR detection of flagellates in mosquito hosts is appreciated. This work was supported by the Czech Grant Agency (P305/11/2179, 206/09/H026) and a Praemium Academiae award to J.L., who is also a Fellow of the Canadian Institute for Advanced Research (CIFAR). Research in the Archibald Lab was supported by an operating grant from the Canadian Institutes of Health Research (CIHR). J.M.A. holds a CIHR New Investigator Award and is a CIFAR Fellow. M.D.L., A.A.P., A.S.K. were supported by Russian Ministry of Education and Science (G34.31.0008). The access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the program LM2010005, is appreciated. The access to the CERIT-SC computing and storage facilities provided under the program Center CERIT Scientific Cloud, part of the Operational Program Research and Development for Innovations, regulation number CZ. 1.05/3.2.00/08.0144, is acknowledged.

Received: January 15, 2013
Revised: June 17, 2013
Accepted: July 12, 2013
Published: September 5, 2013

References

1. Maslov, D.A., Votýpka, J., Yurchenko, V., and Lukeš, J. (2013). Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. Trends Parasitol. *29*, 43–52.
2. Simpson, A.G., Stevens, J.R., and Lukeš, J. (2006). The evolution and diversity of kinetoplastid flagellates. Trends Parasitol. *22*, 168–174.
3. Adams, E.R., Hamilton, P.B., and Gibson, W.C. (2010). African trypanosomes: celebrating diversity. Trends Parasitol. *26*, 324–328.
4. Votýpka, J., Szabová, J., Rádrová, J., Zídková, L., and Svobodová, M. (2012). *Trypanosoma culicavium* sp. nov., an avian trypanosome transmitted by *Culex* mosquitoes. Int. J. Syst. Evol. Microbiol. *62*, 745–754.
5. Van Dyken, M., Bolling, B.G., Moore, C.G., Blair, C.D., Beaty, B.J., Black, W.C., 4th, and Foy, B.D. (2006). Molecular evidence for trypanosomatids in *Culex* mosquitoes collected during a West Nile virus survey. Int. J. Parasitol. *36*, 1015–1023.
6. Pruesse, E., Peplies, J., and Glöckner, F.O. (2012). SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics *28*, 1823–1829.
7. Leonard, G., Soanes, D.M., and Stevens, J.R. (2011). Resolving the question of trypanosome monophyly: a comparative genomics approach using whole genome data sets with low taxon sampling. Infect. Genet. Evol. *11*, 955–959.
8. Votýpka, J., Klepetková, H., Jirků, M., Kment, P., and Lukeš, J. (2012). Phylogenetic relationships of trypanosomatids parasitising true bugs (Insecta: Heteroptera) in sub-Saharan Africa. Int. J. Parasitol. *42*, 489–500.
9. Borghesan, T.C., Ferreira, R.C., Takata, C.S., Campaner, M., Borda, C.C., Paiva, F., Milder, R.V., Teixeira, M.M., and Camargo, E.P. (2013). Molecular phylogenetic redefinition of *Herpetomonas* (Kinetoplastea, Trypanosomatidae), a genus of insect parasites associated with flies. Protist *164*, 129–152.
10. Teixeira, M.M., Borghesan, T.C., Ferreira, R.C., Santos, M.A., Takata, C.S., Campaner, M., Nunes, V.L., Milder, R.V., de Souza, W., and Camargo, E.P. (2011). 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 *162*, 503–524.
11. Jirků, M., Yurchenko, V.Y., Lukeš, J., and Maslov, D.A. (2012). New species of insect trypanosomatids from Costa Rica and the proposal for a new subfamily within the Trypanosomatidae. J. Eukaryot. Microbiol. *59*, 537–547.
12. Jackson, A.P., Quail, M.A., and Berriman, M. (2008). Insights into the genome sequence of a free-living Kinetoplastid: *Bodo saltans* (Kinetoplastida: Euglenozoa). BMC Genomics *9*, 594.
13. Lartillot, N., Brinkmann, H., and Philippe, H. (2007). Suppression of long-branch attraction artefacts in the animal phylogeny using a site-heterogeneous model. BMC Evol. Biol. *7*(*Suppl 1*), S4.
14. Moreira, D., López-García, P., and Vickerman, K. (2004). An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. Int. J. Syst. Evol. Microbiol. *54*, 1861–1875.
15. Deschamps, P., Lara, E., Marande, W., López-García, P., Ekelund, F., and Moreira, D. (2011). Phylogenomic analysis of kinetoplastids supports that trypanosomatids arose from within bodonids. Mol. Biol. Evol. *28*, 53–58.
16. Léger, L. (1904). Sur les affinités de l'Herpetomonas subulata et la phylogénie des trypanosomes. Comptes Rendus des Séances de la Société de Biologie *57*, 615–617.
17. Minchin, E.A. (1908). Investigations on the development of trypanosomes in the tsetse flies and other Diptera. Q. J. Microsc. Sci. *52*, 159–260.

18. Votýpka, J., Maslov, D.A., Yurchenko, V., Jirků, M., Kment, P., Lun, Z.R., and Lukeš, J. (2010). Probing into the diversity of trypanosomatid flagellates parasitizing insect hosts in South-West China reveals both endemism and global dispersal. Mol. Phylogenet. Evol. *54*, 243–253.
19. Týč, J., Votýpka, J., Klepetková, H., Šuláková, H., Jirků, M., and Lukeš, J. (2013). Growing diversity of trypanosomatid parasites of flies (Diptera: Brachycera): frequent cosmopolitism and moderate host specificity. Mol. Phylogenet. Evol. *69*, 255–264.
20. Poinar, G., Jr., and Poinar, R. (2004). *Paleoleishmania proterus* n. gen., n. sp., (Trypanosomatidae: Kinetoplastida) from Cretaceous Burmese amber. Protist *155*, 305–310.
21. Bruce, D. (1895). Preliminary Report on the Tsetse Fly Disease or Nagana, in Zululand (London: Durban, Bennett & Davis).
22. Hughes, A.L., and Piontkivska, H. (2003). Phylogeny of Trypanosomatidae and Bodonidae (Kinetoplastida) based on 18S rRNA: evidence for paraphyly of *Trypanosoma* and six other genera. Mol. Biol. Evol. *20*, 644–652.
23. Runckel, C., Flenniken, M.L., Engel, J.C., Ruby, J.G., Ganem, D., Andino, R., and DeRisi, J.L. (2011). Temporal analysis of the honey bee microbiome reveals four novel viruses and seasonal prevalence of known viruses, *Nosema*, and *Crithidia*. PLoS ONE *6*, e20656.
24. Talavera, G., and Castresana, J. (2007). Improvement of phylogenies after removing divergent and ambiguously aligned blocks from protein sequence alignments. Syst. Biol. *56*, 564–577.
25. Keane, T.M., Creevey, C.J., Pentony, M.M., Naughton, T.J., and McInerney, J.O. (2006). Assessment of methods for amino acid matrix selection and their use on empirical data shows that ad hoc assumptions for choice of matrix are not justified. BMC Evol. Biol. *6*, 29.
26. Shimodaira, H. (2002). An approximately unbiased test of phylogenetic tree selection. Syst. Biol. *51*, 492–508.

Current Biology, Volume *23*

Supplemental Information

Paratrypanosoma Is a Novel

Early-Branching Trypanosomatid

Pavel Flegontov, Jan Votýpka, Tomáš Skalický, Maria D. Logacheva, Aleksey A. Penin, Goro Tanifuji, Naoko T. Onodera, Alexey S. Kondrashov, Petr Volf, John M. Archibald, and Julius Lukeš

A

bodonids
Leishmaniinae (Slow-evolving)
Leptomonas collosoma
Sergeia podlipaevi
Angomonas & Strigomonas
Phytomonas
Herpetomonas
Leptomonas jaculum
Blastocrithidia
Trypanosoma
0.02
Paratrypanosoma

B

Parabodo_caudatus
Euglena_gracilis
Trypanoplasma_borreli
Bodo_saltans
Angomonas & Strigomonas
Phytomonas
Blastocrithidia
Herpetomonas
Leptomonas collosoma
Sergeia podlipaevi
Leishmaniinae (Slow-evolving)
Trypanosoma
0.1
Paratrypanosoma

Figure S2. Related to Figure 2

(A) Maximum likelihood phylogeny based on the SSU rRNA gene, constructed under the GTR+Γ model. *P. confusum* is underlined. Clades containing dixenous species are highlighted with thick lines. The scale bar indicates the inferred number of nucleotide substitutions per site. Bootstrap values > 75 % are displayed. (B) Maximum likelihood phylogeny based on the gGAPDH gene, constructed under the LG+Γ model. *P. confusum* is underlined. Clades containing dixenous species are highlighted with thick lines. The scale bar indicates the inferred number of amino acid substitutions per site. Bootstrap values > 75 % are displayed.

Table S1. Related to Figure 3

Part A. Genomic data sources.

species	data type	proteome size / ORF number	data source	data deposition
Naegleria gruberi (Excavata, Discoba, Discicristata, Heterolobosea)	annotated genome	15,762	Fritz-Laylin et al. 2010	GenBank
Perkinsela sp. CCAP1560/4 (Excavata, Discoba, Discicristata, Euglenozoa, Kinetoplastea, Prokinetoplastina)	transcriptome assembly, ORFs[b]	79,957	unpublished data (G.T., P.F., N.O., J.L., J.A.)	GenBank Acc. No. KC534504-KC534632[a]
Bodo saltans Lake Konstanz (Excavata, Discoba, Discicristata, Euglenozoa, Kinetoplastea, Metakinetoplastina, Eubodonida)	annotated genome	18,963	http://www.genedb.org/Homepage/Bsaltans	GeneDB
Paratrypanosoma confusum (Excavata, Discoba, Discicristata, Euglenozoa, Kinetoplastea, Metakinetoplastina, Trypanosomatida)	draft genome assembly, ORFs[c]	41,836	this study	GenBank Acc. No. KC534633-KC534828[a]
Trypanosoma cruzi CL Brenner, non-Esmeraldo chromosomes	annotated genome	9,386	El-Sayed et al. 2005; Weatherly et al. 2009	TriTrypDB v. 4.0
Trypanosoma cruzi CL Brenner, Esmeraldo chromosomes	annotated genome	9,042	El-Sayed et al. 2005; Weatherly et al. 2009	TriTrypDB v. 4.0
Trypanosoma vivax Y486	annotated genome	11,870	http://www.genedb.org/Homepage/Tvivax	TriTrypDB v. 4.0
Trypanosoma congolense IL3000	annotated genome	13,459	http://www.genedb.org/Homepage/Tcongolense	TriTrypDB v. 4.0
Trypanosoma brucei brucei TREU 927	annotated genome	9,274	Berriman et al. 2005	TriTrypDB v. 4.0
Trypanosoma brucei gambiense DAL 972	annotated genome	9,844	Jackson et al. 2010	TriTrypDB v. 4.0
Phytomonas serpens 9T	draft genome assembly, ORFs[c]	19,488	Kořený et al. 2012	NCBI Acc. No. PRJNA80957

Crithidia fasciculata Cf-C1	draft genome assembly, ORFs[c]	49,110	unpublished data (Beverley S.M. and The Genome Center, Washington University School of Medicine)	TriTrypDB v. 4.0
Leptomonas pyrrhocoris H10	draft genome assembly, ORFs[c]	37,299	unpublished data (P.F., M.D.L., A.A.P., A.S.K., J.L.)	GenBank Acc. No. KC543586-KC543700[a]
Endotrypanum monterogeii LV88	draft genome assembly, ORFs[c]	38,570	unpublished data (Beverley S.M. and The Genome Center, Washington University School of Medicine)	TriTrypDB v. 4.0
Leishmania braziliensis MHOM/BR/75/M2904	annotated genome	8,201	Peacock et al. 2007	TriTrypDB v. 4.0
Leishmania tarentolae Parrot-TarII	draft genome assembly, ORFs[c,d]	49,916	Raymond et al. 2012 (annotated genome)	TriTrypDB v. 4.0
Leishmania mexicana MHOM/GT/2001/U1103	annotated genome	8,151	http://www.genedb.org/Homepage/L mexicana	TriTrypDB v. 4.0
Leishmania infantum MCAN/ES/98/LLM-877	annotated genome	8,187	Peacock et al. 2007	TriTrypDB v. 4.0
Leishmania major MHOM/IL/80/Friedlin	annotated genome	8,319	Ivens et al. 2005	TriTrypDB v. 4.0

[a] Accession numbers for ORF sequences used in this study.

[b] ORFs > 100 amino acids in length from stop to stop codons were extracted.

[c] ORFs > 100 amino acids in length from AUG to stop codons were extracted.

[d] The un-annotated version of the genome was used in this study.

Part B. Concatenated multi-protein alignments used in this study.

outgroup	alignment identity cut-off, % [a]	number of sequence clusters	gaps removed [b]	aligned amino acids	variable positions
Naegleria gruberi	40 [c]	114	yes	26,058	16,049
			no	36,422	22,628
	50 [d]	52	yes	12,607	6,838
			no	17,140	9,423
	no [d]	226	yes	52,678	37,547
			no	74,640	53,024
Perkinsela sp.	40 [c]	129	yes	32,400	20,448
			no	45,166	28,501
Naegleria gruberi, *Perkinsela* sp.	40 [d]	42	yes	10,602	5,515
			no	13,654	7,799

[a] Average alignment identity in the BLASTP hit region between the outgroup sequence and the other sequences was calculated.

[b] Positions with gaps were removed altogether, or positions with gaps in <50% of the sequences were kept.

[c] Principal datasets.

[d] Auxiliary datasets.

Table S2. Related to Figure 3

Part A. Results of maximum likelihood (ML) and Bayesian phylogenetic analyses with concatenated multi-protein alignments.

				ML phylogeny (RAxML 7.2.8)			Bayesian phylogeny (PhyloBayes 3.3f)				
outgroup	**alignment identity cut-off, %**	**number of sequence clusters**	**gaps removed**	**phylogenetic model**	**lnL for best topology**	**bootstrap support for *P. confusum* branch**	**phylogenetic model**	**posterior probability for *P. confusum* branch**	**maximum difference in bipartition frequencies**	**conflicting bipartitions [a]**	**Model cross-validation scores with standard deviation, Poisson+Γ+CAT vs. GTR+Γ+CAT**
Naegleria gruberi	40	114	yes	LG+Γ+F	-304,446	100	Poisson+Γ+CAT	0.97	0.110	*N. gruberi, P. confusum;* *P. confusum, B. saltans*	-361.96 +/- 39.41
				GTR+Γ	-303,066	100	GTR+Γ+CAT	1	0	N/A	
			no	LG+Γ+F	-423,734	100	Poisson+Γ+CAT	0.99	0.032	*N. gruberi, P. confusum*	N/A
				GTR+Γ	-421,801	100	GTR+Γ+CAT	1	0	N/A	
	50	52	yes	LG+Γ+F	-123,918	100	Poisson+Γ+CAT	0.98	0.037	*P. confusum, B. saltans;* *N. gruberi, P. confusum*	-144.44 +/- 16.03
				GTR+Γ	-123,186	100	GTR+Γ+CAT	1	0	N/A	
			no	LG+Γ+F	-171,869	100	Poisson+Γ+CAT	0.99	0.037	*N. gruberi, P. confusum;* *P. confusum, B. saltans*	-186.76 +/- 31.08
				GTR+Γ	-170,882	100	GTR+Γ+CAT	1	0	N/A	
	no	226	yes	LG+Γ+F	-733,362	100	Poisson+Γ+CAT	0.74	0.318	*P. confusum, B. saltans;* *N. gruberi, P.*	N/A

										confusum	
			no	LG+Γ+F	-1,011,826	100	Poisson+Γ+CAT	0.59	0.361	*N. gruberi, P. confusum; P. confusum, B. saltans*	
Perkinsela sp.	40	129	yes	LG+Γ+F	-391,850	100	Poisson+Γ+CAT	1	0	N/A	-446.25 +/- 36.69
				GTR+Γ	-390,134	100	GTR+Γ+CAT	1	0	N/A	
			no	LG+Γ+F	-545,644	100	Poisson+Γ+CAT	1	0	N/A	N/A
				GTR+Γ	-543,262	100	GTR+Γ+CAT	1	0	N/A	
Naegleria gruberi, Perkinsela sp.	40	42	yes	LG+Γ+F	-132,144	100	Poisson+Γ+CAT	0.99	0.018	*P. confusum, B. saltans*	-154.01 +/- 16.32
				GTR+Γ	-131,534	100	GTR+Γ+CAT	1	0	N/A	
			no	LG+Γ+F	-169,941	100	Poisson+Γ+CAT	0.95	0.065	*P. confusum, B. saltans*; *N. gruberi, Perkinsela* sp., *P. confusum*	N/A
				GTR+Γ	-169,189	100	GTR+Γ+CAT	1	0	N/A	

[a] Conflicting bipartitions, separated by a semicolon, are listed in the order of decreasing probability.

Part B. Posterior predictive test of mutational saturation in PhyloBayes 3.3f under the Poisson+Γ+CAT phylogenetic model.

outgroup	alignment identity cut-off, %	number of sequence clusters	gaps removed	average observed number of substitutions per site	average posterior predictive number of substitutions per site	P-value	average observed number of homoplasies per site	average posterior predictive number of homoplasies per site	P-value
Naegleria gruberi	40	114	yes	9.41	9.42	0.54	5.70	5.71	0.66
			no	10.91	10.92	0.56	6.96	6.97	0.68
	50	52	yes	5.96	5.96	0.56	3.07	3.08	0.67
			no	6.79	6.79	0.56	3.71	3.72	0.67
	no	226	yes	18.29	18.29	0.54	13.01	13.02	0.66
			no	22.98	22.98	0.56	17.22	17.23	0.68
Perkinsela sp.	40	129	yes	10.11	10.12	0.55	6.23	6.25	0.69
			no	10.73	10.73	0.57	6.74	6.75	0.72

Table S3. Single-Gene Tree Topologies under the LG+Γ Model, Related to the Results

outgroup		*Naegleria gruberi*		***Perkinsela* sp.**
alignment identity cut-off, %	40	50	no	40
number of sequences	114	52	226	129
gaps removed	no	no	no	no
topologies:				
outgroup, (*B. saltans*, (*P. confusum*, (trypanosomatids)))	29	12	54	36
(*B saltans* + *P. confusum*) [a,b]	25	9	65	33
P. confusum, (*Trypanosoma*) [a]	15	8	25	17
outgroup, (*P. confusum*, (*B. saltans*, (trypanosomatids))) [c]	9	5	19	8
P. confusum, (*Phytomonas*, (Leishmaniinae)) [a]	7	2	13	8

[a] The exclusive species group may be located anywhere on the tree.

[b] The topologies likely resulting from long-branch attraction (LBA).

Supplemental References

1. Berriman, M., Ghedin, E., Hertz-Fowler, C., Blandin, G., Renauld, H., Bartholomeu, D.C., et al. (2005). The genome of the African trypanosome *Trypanosoma brucei*. Science. *309*, 416-422.

2. El-Sayed, N.M., Myler, P.J., Bartholomeu, D.C., et al. (103 co-authors). (2005). The genome sequence of *Trypanosoma cruzi*, etiologic agent of Chagas disease. Science. *309*, 409-415.

3. Ivens, A.C., Peacock, C.S., Worthey, E.A., et al. (101 co-authors). (2005). The genome of the kinetoplastid parasite, *Leishmania major*. Science. *309*, 436-442.

4. Jackson, A.P., Sanders, M., Berry, A., et al. (14 co-authors). (2010). The genome sequence of *Trypanosoma brucei gambiense*, causative agent of chronic human african trypanosomiasis. PLoS Negl. Trop. Dis. *4*, e658.

5. Kořený, L., Sobotka, R., Kovářová, J., Gnipová, A., Flegontov, P., Horváth, A., Oborník, M., Ayala, F.J., and Lukeš, J. (2012). Aerobic kinetoplastid flagellate *Phytomonas* does not require heme for viability. Proc. Natl. Acad. Sci. USA. *109*, 3808-3813.

6. Peacock, C.S., Seeger, K., Harris, D., et al. (42 co-authors). (2007). Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. Nat. Genet. *39*, 839-847.

7. Raymond, F., Boisvert, S., Roy, G., et al. (12 co-authors). (2012). Genome sequencing of the lizard parasite *Leishmania tarentolae* reveals loss of genes associated to the intracellular stage of human pathogenic species. Nucleic Acids Res. *40*, 1131-1147.

8. Weatherly, D.B., Boehlke, C., and Tarleton, R.L. (2009). Chromosome level assembly of the hybrid *Trypanosoma cruzi* genome. BMC Genomics. *10*, 255.

5.2 Paper II

Lukeš J, **Skalický T**, Týč J, Votýpka J, Yurchenko V (2014) Evolution of parasitism in kinetoplastid flagellates. **Molecular and Biochemical Parasitology** 195(2):115-122 (IF = 2.068).

Abstract

Kinetoplastid protists offer a unique opportunity for studying the evolution of parasitism. While all their close relatives are either photo- or phagotrophic, a number of kinetoplastid species are facultative or obligatory parasites, supporting a hypothesis that parasitism has merged within this group of flagellates. In this review, we discuss origin and evolution of parasitism in bodonids and trypanosomatids and specific adaptations allowing these protozoa to co-exist with their hosts. We also explore the limits of biodiversity of monoxenous (one host) trypanosomatids and some features distinguishing them from their dixenous (two hosts) relatives.

Molecular & Biochemical Parasitology 195 (2014) 115–122

Contents lists available at ScienceDirect

Molecular & Biochemical Parasitology

Review

Evolution of parasitism in kinetoplastid flagellates

Julius Lukeš [a,b,*], Tomáš Skalický [a,b], Jiří Týč [a,b], Jan Votýpka [a,c], Vyacheslav Yurchenko [a,d]

[a] Biology Centre, Institute of Parasitology, Czech Academy of Sciences, Czech Republic
[b] Faculty of Science, University of South Bohemia, České Budějovice (Budweis), Czech Republic
[c] Department of Parasitology, Faculty of Sciences, Charles University, Prague, Czech Republic
[d] Life Science Research Centre, Faculty of Science, University of Ostrava, Ostrava, Czech Republic

ARTICLE INFO

Article history:
Available online 2 June 2014

Keywords:
Evolution
Phylogeny
Vectors
Diversity
Parasitism
Trypanosoma

ABSTRACT

Kinetoplastid protists offer a unique opportunity for studying the evolution of parasitism. While all their close relatives are either photo- or phagotrophic, a number of kinetoplastid species are facultative or obligatory parasites, supporting a hypothesis that parasitism has emerged within this group of flagellates. In this review we discuss origin and evolution of parasitism in bodonids and trypanosomatids and specific adaptations allowing these protozoa to co-exist with their hosts. We also explore the limits of biodiversity of monoxenous (one host) trypanosomatids and some features distinguishing them from their dixenous (two hosts) relatives.

© 2014 Elsevier B.V. All rights reserved.

Contents

1. Emergence of parasitism: setting (up) the stage 115
2. Diversity *versus* taxonomy: closing the gap 116
3. Diversity is not limitless: defining its extent 117
4. Acquisition of parasitic life style: the "big" transition 118
5. Acquisition of dixenous life style: the "small" transition 119
6. Monoxenous life style: back again 120
7. Adaptations to parasitism: are there any? 120
Acknowledgements 121
References 121

1. Emergence of parasitism: setting (up) the stage

With a certain degree of simplification, when the frequency of eukaryotic parasites encountered in vertebrate and invertebrate hosts is considered, probably only apicomplexans surpass kinetoplastid protists in abundance and diversity, and only parasitic nematodes seem to have a broader host range [1,2]. Kinetoplastids are evolutionarily more ancestral compared to the majority of other groups of parasitic protists, widespread and adaptable, which is an apparent reflection of their extremely successful life style. A recent taxonomy places Kinetoplastea along with its three sister groups (Euglenida, Symbiontida and Diplonemea) into Euglenozoa that belongs to the Discicristata, a group of protists unified by a striking feature–discoidal mitochondrial cristae [3] (Fig. 1). Euglenida are phototrophic or less frequently phagotrophic, the latter life strategy being characteristic for all known symbiontids and diplonemids [3]. Accordingly, parasitism must have emerged uniquely in the kinetoplastid lineage. It is an exciting challenge to identify genetic changes and/or inventions underlying this dramatic switch to a parasitic life style; however, it has to be postponed until the whole genomes for these sister clades of kinetoplastids are available.

Phylogenetic evidence strongly supports the early-branching of Prokinetoplastina within Kinetoplastea. This tiny group harbors only two known representatives – *Ichthyobodo* and *Perkinsela* (Fig. 2) [4,5]. While *Ichthyobodo* (also called *Costia*) is a bi-flagellar ectoparasite of fish, *Perkinsela* (also known as PLO, parasome and *Perkinsiella*) resides directly in the cytoplasm of certain amoebae parasitizing the gills of fish. This aflagellar kinetoplastid seems

* Corresponding author at: Biology Centre, Institute of Parasitology, Czech Academy of Sciences, Branišovská 31, 370 05, České Budějovice (Budweis), Czech Republic. Tel.: +420 385310351; fax: +420 385310388.
E-mail address: jula@paru.cas.cz (J. Lukeš).

http://dx.doi.org/10.1016/j.molbiopara.2014.05.007
0166-6851/© 2014 Elsevier B.V. All rights reserved.

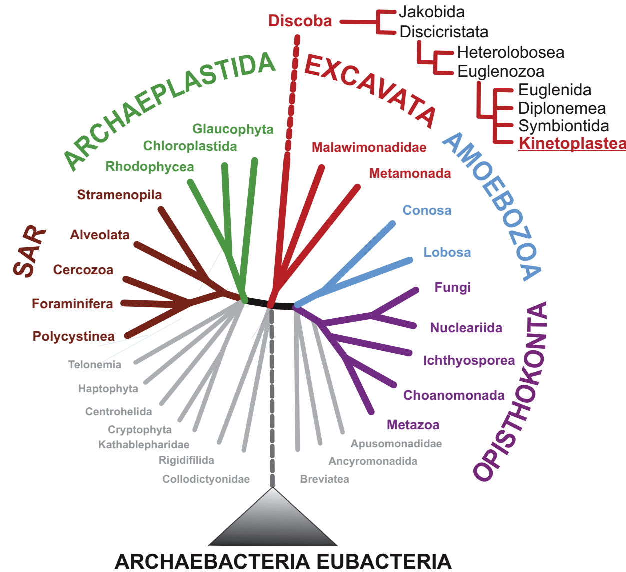


Fig. 1. A view of general eukaryote phylogeny reflecting the recent classification (based on [3]) and highlighting the taxonomic position of Kinetoplastea.

to behave like an organelle, invariably located close to the host nucleus and dividing synchronously with the host cell [6]. Based on DAPI staining, mitochondrial (=kinetoplast [k]) DNA of *Perkinsela* seems to be much more abundant than its nuclear DNA [7]. It will be exciting to investigate whether the extremely tight relationship with the amoeba host is reflected in the kDNA and nuclear genome of *Perkinsela*. Due to its robust branching at the basis of the Kinetoplastea clade, it is tempting to interpret the endosymbiont-like intracellularity of *Perkinsela* as some ancestral form of parasitism *via* which the kinetoplastid invaded first hosts. However, the absence of flagella, which are otherwise present in all sister clades (euglenids, symbiontids and diplonemids) as well as in all derived lineages, qualifies *Perkinsela* as a unique case of parasitic reductionism.

All the remaining bodonids fall into Metakinetoplastina, a group further subdivided into four clades (Neobodonida, Parabodonida, Eubodonida and Trypanosomatida) (Fig. 2), of which only the latter is obligatory parasitic [3,4]. Mutual relationships within the bodonids are far from being firmly established, yet it is obvious that they acquired parasitic life style independently more than once. Still, only a handful of parasitic bodonids is known, whereas some free-living species are virtually omnipresent and ecologically highly significant [5,8]. Members of the genera *Trypanoplasma* and *Cryptobia* parasitize fish and snails [9,10], respectively. *Azumiobodo hoyamushi* causes economically important damage to cultured ascidians [11], while *Jarrellia attramenti* found in the blowhole of whales and dolphins [12] may rather be a commensal than a parasite (Fig. 2). For the purpose of this review, we will focus on flagellates belonging to Trypanosomatida as they embrace an absolute majority of parasitic species (see below).

2. Diversity *versus* taxonomy: closing the gap

The taxonomy of Trypanosomatida was originally defined by a set of morphotypes, which differ in respect to the mutual positions of the kDNA, nucleus and flagellar pocket, and the presence or derived loss of a single flagellum [13–16]. Extensive application of electron microscopy in studies of trypanosomatids did not add any important distinguishing features [17,18]. Since the advent of molecular methods it became obvious that neither the individual morphotypes nor their combination within a given life cycle hold any taxonomic value, as they are randomly distributed in the sequence-based phylogenetic trees [19]. Moreover, it seems plausible that there is a continuum of cell forms rather than eight distinct morphotypes.

Due to this dearth of morphological features, one has to resort to DNA sequencing in order to establish taxonomic position of a given trypanosomatid flagellate. There are two categories of genes of choice suitable for this purpose: the small subunit (SSU) rRNA and the glycosomal glyceraldehyde-3-phosphate dehydrogenase (gGAPDH) genes are informative for higher level taxonomy, and are usually sufficient for the genus-level ranking [20–22], while the sequences of the spliced leader (SL) RNA gene and the respective intergenic region allow distinguishing among individual species or even populations [23–27]. The growing number of species and strains, for which sequence data are available, revealed the artificial character of all previously described monoxenous (=one invertebrate host) genera, however, all three dixenous (=vertebrate or plant host and invertebrate vector) genera *Trypanosoma*, *Leishmania* and *Phytomonas* remain monophyletic and well supported (Fig. 3) [19,28].

One approach to close the gap between the outdated morphology-based taxonomy and the molecular-based cladistics that better reflects the relationships among trypanosomatids is to attach taxonomic units to the latter clades. Using this approach, some decades-old taxa rendered paraphyletic by molecular studies and hence invalidated, can be "recycled", *i.e.* used just for a single clade containing the type species of a given genus. This solution is taxonomically acceptable, and was successfully used in several instances so far [29–31]. In an alternative approach, novel clades

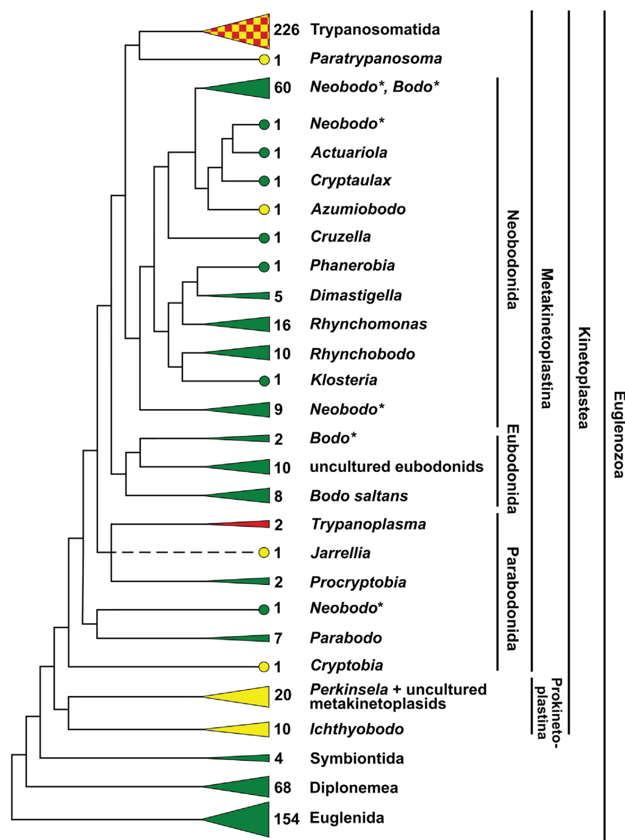


Fig. 2. Evolutionary relationships among bodonids based on SSU rRNA sequences. Numbers of available sequences representing individual species are shown as well as the new high level classification. Circles and triangles denote single and multiple known representatives of a particular clade, respectively. Dotted line indicates morphology-based position in the absence of sequence data. Green color depicts free-living species; yellow and red colors represent monoxenous and dixenous parasites, respectively. The intentionally collapsed clade of trypanosomatid flagellates (Trypanosomatida) is highlighted by a yellow-red checkerboard. The paraphyletic genera *Neobodo* and *Bodo* are labeled with asterisks.

that are consistently and highly supported in the phylogenetic reconstructions and contain at least several isolates are labeled with new generic names [32]. Based on the available sequence data, one can predict the existence of no more than two dozens of clades deserving the level of a genus, making a combination of these approaches not only feasible but also practical. The presently recognized taxonomic units are summarized in Fig. 3.

3. Diversity is not limitless: defining its extent

Defining the higher level taxonomy of Kinetoplastea is important for matching the new molecular data with a taxonomic framework constructed for over 100 years, but unfortunately it tells very little about the true diversity of these widespread protists. Bodonid SSU rRNA sequences obtained mostly in frame of environmental studies are summarized in Fig. 2. These bi-flagellar kinetoplastids are present in all aquatic ecosystems, yet they rarely abound and hence tend to be ignored. In the absence of extensive morphological analysis of bodonids it remains to be established whether the present taxonomy is robust or artificial. Some bodonid species are globally distributed [33], which supports the “everything is everywhere” paradigm formulated for the free-living protists by Fenchel and Finlay [34]. However, while morphological analyses seemingly support this hypothesis, we do not have enough genetic data yet to confirm it properly because many various isolates are indistinguishable from each other by morphology but significantly different in genetic analyses [8,33].

Much more data is available for the obligatory parasitic trypanosomatids, especially members of the dixenous genera. Due to the medical relevance of trypanosomes and leishmanias as causative agents of sleeping sickness, Chagas disease and leishmaniases, Trypanosomatida has been attracting most of the researchers' attention. Hence, at least some sequence information is available for over 100 *Trypanosoma* and 35 *Leishmania* species in the NCBI database. This allows their straightforward detection and determination, which is needed given the pathogenicity of many of them for humans and economically important vertebrates. Whole genomes have been sequenced for several strains/subspecies of

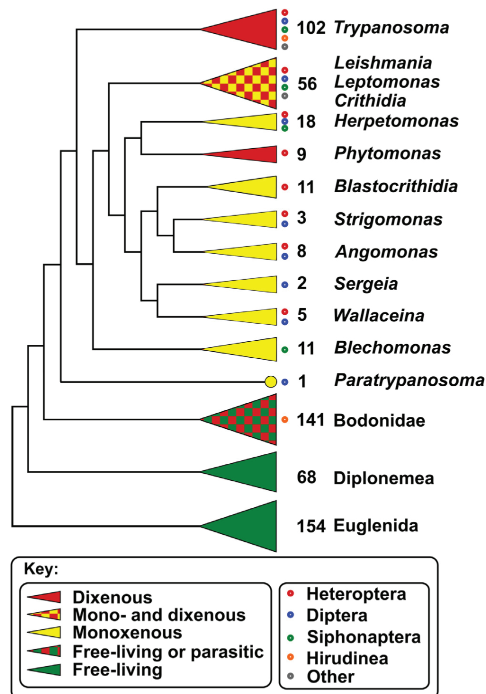


Fig. 3. Evolutionary relationships among trypanosomatids based on SSU rRNA sequences. Numbers of available sequences representing individual species are shown. Green color depicts free-living outgroups, the intentionally collapsed clade of free-living and parasitic bodonid flagellates (Bodonidae) is highlighted by a red-green checkerboard. Yellow and red colors represent monoxenous and dixenous parasites, respectively. The yellow-red checkerboard depicts a clade with mixed life cycles. Small colored circles depict major groups of invertebrate hosts.

T. brucei, *T. cruzi*, *T. congolense*, *T. vivax*, *L. major*, *L. donovani*, *L. infantum*, *L. braziliensis* and *L. mexicana* (www. http://tritrypdb.org/), and most recently also for two species of *Phytomonas* [35], with many more being in the pipeline. Due to the lack of economic significance, the monoxenous trypanosomatids of insects have been largely overlooked, with most species being described solely on the basis of their morphology and host specificity [36,37]. This has changed recently, and the SL and SSU rRNA sequences have been deposited for almost 150 species. Moreover, a draft-quality assembly of the whole genome of at least one monoxenous species, *Angomonas deanei* [38], and the unassembled reads of *Leptomonas seymouri* and the early model trypanosomatid *Crithidia fasciculata* are now available (http://www.sanger.ac.uk/resources/downloads/protozoa/).

The total number of extant monoxenous trypanosomatids might be staggeringly high given the extreme species richness of their insect hosts [39]. In order to tackle this potentially enormous landscape, we and our collaborators have recently established a system based on the so-called typing units (TUs), defined on the basis of >10% sequence divergence of the SL RNA gene [19,23,24,26,27,31]. A publicly available database which would contain the information of the host, location of infection, geographical origin, date of isolation, availability in culture *etc.* for each TU, is under preparation (J.V., J.L., V.Y. and D. A. Maslov, unpubl. data). We encourage the community to deposit information on all trypanosomatids isolated from insect and other invertebrate hosts into this shared database. Although much additional data is needed for more definitive conclusions, it is already obvious that several TUs (and likely species) are globally distributed [26], and some TUs are confined to a (very) narrow host range, while others are opportunists [24,25,29,32,40,41]. Now we know several heteropterid, siphonapterid and/or dipterid species that host more than a single TU, but it also appears that many insect orders are never or only rarely infected with trypanosomatids [19].

4. Acquisition of parasitic life style: the “big” transition

Species-rich and morphologically diverse euglenids and diplonemids are almost exclusively free-living, so this life style was likely the ancestral state of early kinetoplastids. The tree in Fig. 2 shows the generally accepted paraphyly of bodonids, and independent establishment of parasitic or commensalic life styles in five to six lineages (*Azumiobodo*, *Cryptobia*, *Ichthyobodo*, *Jarrellia*, *Perkinsela* and *Trypanoplasma*). Interestingly, with the sole exception of *Cryptobia helicis* [42], all parasitic bodonids remain confined to aquatic hosts.

Morphological and molecular evidence strongly support the origin of obligatory parasitic trypanosomatids from the clade of free-living *Bodo saltans* [5,43–45]. Until recently, the most basal branch of this clade was the genus *Trypanosoma*, supporting the scenario in which ancestral trypanosomatids established themselves in vertebrates already at an early stage, and only subsequently invaded insects *via* their blood-sucking fellows [46]. Such chain of events was challenged by fossil evidence in the form of a flagellate morphologically indistinguishable from the extant *Leishmania*, discovered in an amber-trapped blood-engorged mosquito (Fig. 4) [47,48]. This so-called *Paleoleishmania* dated to approximately 150 MYA was found among nucleated erythrocytes, prompting speculations that these originated from a dinosaur [47]. If correct, this finding would place the invasion of vertebrates by trypanosomatids to the early Cretaceous period at the latest. Phylogeographic data must be taken with caution as has been recently exemplified for trypanosomes of South American alligators and African crocodiles. While Crocodylidae and Alligatoridae split in mid Cretaceous (∼90 MYA), their parasites are still closely related testifying to a much later separation due to the marine circumtropical dispersal at the Miocene/Pliocene boundary (∼4–5 MYA) [49]. An ancestor of *Leishmania* probably developed dixenous life cycle in the warm-blooded vertebrates in late Cretaceous about 85 million years ago [50], around the time of divergence of mammalian orders and the first fossil records of its vector (Phebotominae) (Fig. 4). This event likely took place in Neotropics (present South America) in sloths, which have lower body temperature [51–53]. However, an alternative hypothesis places the marsupials in the spotlight: dixenous (*e.g. Trypanosoma cruzi* or *Leishmania* spp.) and monoxenous trypanosomatids of the genera *Crithidia*, *Leptomonas* and *Herpetomonas* can survive and multiply in their anal scent glands. Monoxenous species could pre-adapt there to the dixenous life cycle because within the scent glands parasites are protected from the host immune system and the body temperature is lower [54].

Recent description of the likely monoxenous *Paratrypanosoma confusum*, which constitutes a well-supported branch between the free-living *B. saltans* on one side and *Trypanosoma* plus all other trypanosomatids on the other side [28,45] supports a scenario, in which the ancestral flagellate first invaded insects or other invertebrate hosts and only subsequently, probably by blood feeding, entered vertebrates – a theory proposed by Léger in 1904 [55]. Comparative analysis of the genomes of *B. saltans* [43] and *P. confusum* (T.S. et al., unpubl. data) may shed key light on this dramatic change in life strategy.

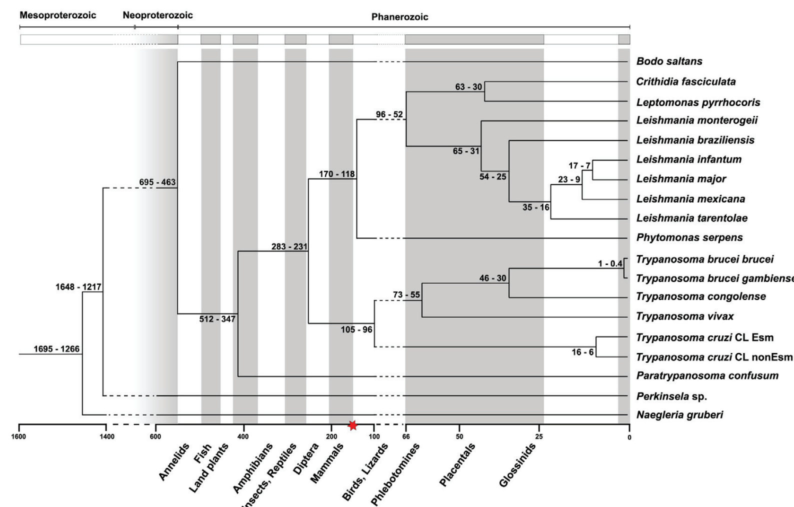


Fig. 4. Phylogenetic tree of kinetoplastid flagellates constructed under the clock model. Concatenated dataset of 42 proteins from 18 kinetoplastid species and *Naegleria* serving as an outgroup described previously, was used for molecular dating [28,98]. Divergence time estimates were inferred by the PhyloBayes 3.3f using CIR process [99] and birth-death prior with implemented soft bounds [100]. A star denotes single known fossil evidence of a kinetoplastid [47]. On the *x* axis absolute time scale in millions of years is shown along with the estimated emergence of host groups parasitized by kinetoplastids [101]. Nodes are at mean divergence and the numbers represent 95% confidence intervals.

Monoxenous trypanosomatids circulate among their insect hosts *via* contamination, coprophagy, necrophagy or predation, with the fecal transmission being probably the predominant way. Transmission among dipteran and hemipteran insects, which host over three-fourths of all described monoxenous species, is favored by their habit of feeding on rich organic sources that are often contaminated by excretions from infected specimens [16,19].

5. Acquisition of dixenous life style: the "small" transition

The origin of the dixenous life style has been discussed several times and the insect-early scenario is now generally favored, since the two-hosts genera *Leishmania* and *Phytomonas* are phylogenetically nested within clades of the single-host trypanosomatids (Fig. 3) [30]. Current phylogenies also almost invariably support a view, in which the derived dixenous life style evolved from the monoxenous one independently for each of the genera *Trypanosoma*, *Leishmania* and *Phytomonas* [19,28,45]. Moreover, it seems that some monoxenous trypanosomatids occasionally attempt a switch to dixeny, or what we call here a "small" transition, as there are several documented cases of mammals and birds being infected with *Herpetomonas megaseliae* and *Crithidia* sp., respectively [16].

More recently, a hypothesis claiming that the monoxenous trypanosomatids may explore new niches particularly in immunocompromised hosts has been put forward [56]. It stemmed from the observation that immunosuppressed HIV-positive patients were often found to host flagellates such as *T. cruzi* or *Leishmania* spp. [57,58], as well as presumably monoxenous species [59]. One such example is a close relative of *Blechomonas* (formerly *Leptomonas*) *pulexsimulantis* isolated from an HIV-positive individual in Brazil [60], which usually parasitizes fleas [32]. An independent line of evidence originated from studies of visceral leishmaniases or kala-azar, caused by the *Leishmania donovani* complex [61]. *Phlebotomus argentipes*, a sand fly vector implicated in the transmission of leishmaniases in Nepal, was estimated to be infected almost as often with the non-*Leishmania* species as with genuine *Leishmania* [62]. The only species recovered so far from co-infections with *Leishmania donovani* is *Leptomonas seymouri* originally isolated from a hemipterid bug [63–65]. The exact ratio between these two parasites is hard to estimate *in situ* because they are morphologically very similar [19,30]. Importantly, *Leptomonas seymouri* outgrows *Leishmania donovani in vitro* resulting in primary cultures enriched with this species [64]. Consequently, when the whole genomes of three strains isolated from kala-azar patients were analyzed, over 95% of sequencing reads belonged to *Leptomonas seymouri* [63]. This also explains another important mystery noted by us (V.Y. and J.V., unpubl. data), namely that several entries reported in GenBank as *Leishmania donovani* are, in fact, sequences of *Leptomonas seymouri*.

We posit that there are several species of monoxenous trypanosomatids capable of surviving within the warm-blooded vertebrate hosts. The evolutionary advantages of such exploration are obvious – potential new ecological niches may suit parasite's needs better and thus facilitate its prevalence and distribution. Nevertheless, everything comes with a price tag. In this particular case it means that parasites must adapt to new environmental conditions within vertebrates which are dramatically different from what they have experienced in insects. Molecular mechanisms of such adaptations are presently unknown. One obvious and critical factor is temperature, which in contrast to invertebrates is constantly high in the warm-blooded vertebrates. Observations discussed above imply that at least some monoxenous trypanosomatids may withstand a thermal shock. This has been proven experimentally for a hemipteran-derived *Crithidia* sp., which was capable of infecting chicken embryos [66]. Two other monoxenous species available in culture (*C. hutneri* and *C. luciliae thermophila*) are known to survive at elevated temperatures [67,68]. Moreover, *Leptomonas seymouri* can also tolerate temperatures of the range observed in

warm-blooded vertebrates (V.Y., J.V., and J.L., in prep.). The molecular nature of adaptations needed for a successful transition from the monoxenous to dixenous life cycle remains to be investigated.

The above-mentioned tendency of monoxenous flagellates to switch for the dixenous life style can be further exemplified by *Blastocrithidia culicis*, which colonizes not only mosquito's digestive tract and haemocoel, but also its salivary glands, raising the possibility of transmission to vertebrates by a mosquito bite [69]. However, there are only very few records of monoxenous species from the blood-sucking vectors of *Trypanosoma* (*e.g.* tsetse flies) and *Leishmania* (sand flies) [5]. The criteria generally used to incriminate vectors are: (i) presence in the same environment as the host; (ii) feeding on the host; (iii) development of the parasite to a transmissible infective stage; (iv) presence of the same parasite in wild-caught vectors and hosts, and finally, (v) transmission *via* a bite, ingestion, or feces. Trypanosomes can be transmitted by saliva (Salivaria), feces (Stercoraria), mechanically, and by ingestion of the vector. Mechanical transmission of some dixenous flagellates by a wide range of vectors has been demonstrated in the laboratory, yet this route does not seem to play a significant role under natural conditions [70]. Leishmanias use several ways of transmission to the vertebrate host: (i) members of the subgenus *Leishmania* are transmitted by inoculation through parasite regurgitation (back flow), although their occurrence in the sand fly salivary glands was also repeatedly reported; (ii) subgenus *Viannia* is transferred similarly, however, prediuresis (excreting urine to concentrate proteins of the bloodmeal while feeding on the host) is also considered as a possible route; (iii) subgenus *Sauroleishmania* is transmitted by defecation, prediuresis or by ingestion of the vector, and finally; (iv) some *Leishmania* (*Endotrypanum*) changes its hosts by contamination [71,72].

Most vectors of the dixenous genera *Trypanosoma*, *Leishmania* and *Phytomonas* are haematophagous or phytophagous members of the orders Hemiptera and Diptera, however other insects or invertebrates such as mites and leeches can also serve as vectors [73–75]. The list of new invertebrate vectors of dixenous species was extended by blood-sucking terrestrial leeches (Haemadipsidae) transmitting a new trypanosome lineage [76], while ticks (*Ixodes* spp.) and midges (Ceratopogonidae) have a potential to transmit *T. copemani* [77] and *Leishmania* spp., respectively [78]. Moreover, fleas (Siphonaptera) were recently shown to be surprisingly frequent hosts of a new clade of monoxenous trypanosomatids named *Blechomonas*. These holometabolous insects undergo a radical metamorphosis and since the only meal of adult fleas is blood, the infections must be established already at the larval stage, with the parasites accompanying the hosts throughout their metamorphosis [32]. In addition, free-living ciliates were occasionally shown to harbor trypanosomatids, potentially serving as their reservoirs [7]. Probably because of the long evolutionary association, the majority of trypanosomes have developed well-balanced relationships with their invertebrate hosts that allow mutual survival, although in some cases virulence was retained (*e.g. T. rangeli* and subspecies of *T. brucei*). Similar logic applies to the relationship between trypanosomes and their vertebrate hosts. Most species cause mild infections with no obvious symptoms, yet for very good reasons, most attention is given primarily to those highly pathogenic for humans and economically important mammals [79].

Dixenous hemoflagellates adapted their transmission to ecological conditions and environment of their hosts. Although *Trypanoplasma* and *Trypanosoma* are unrelated, the former being a bodonid and the latter a trypanosomatid, they both circulate in the blood of marine and freshwater fishes and both are transmitted exclusively by leeches [10,73]. Same principle applies to different avian trypanosomes, which use for their transmission a wide range of dipteran insects commonly attacking birds, such as black flies (Simuliidae), flat flies (Hippoboscidae), mosquitoes (Culicidae) and biting midges (Ceratopogonidae) [80,81]. Similarly, trypanosomes parasitizing bats are transmitted by blood-sucking bugs (Cimicidae and Triatominae) and/or sand flies (Phlebotominae) living within or close to their host's colonies. Infections usually occur when a grooming bat ingests an infected insect or its feces [82]. Other mammalian trypanosomes are transmitted by blood-sucking horse flies (Tabanidae) and sheep keds (Hippoboscidae) and bugs (Triatominae), as well as by fleas (Siphonaptera), which repeatedly defecate on the skin and fur while taking a blood meal [83].

6. Monoxenous life style: back again

The capacity for cyclic development of the Salivarian trypanosomes (species that terminate development in the salivary glands) depends on the suitability of available vectors. *Trypanosoma congolense*, *T. brucei gambiense*, *T. b. rhodesiense* and *T. b. brucei* are (fortunately) confined to the tsetse belt in sub-Saharan Africa, since no other vector than tsetse flies (Glossinidae) is capable of their transmission [84,85]. Two other members of the *T. brucei* species complex escaped the tsetse belt: *T.* (*b.*) *evansi* has adopted horse flies (Tabanidae), other insects (Stomoxyinae and Hippoboscidae) and even vampire bat (*Desmodus* spp.) as its apparently surrogate vectors, while *T.* (*b.*) *equiperdum* has avoided using vectors completely and switched to direct sexual transmission [86,87].

7. Adaptations to parasitism: are there any?

When applied to trypanosomatid parasites, the conventional thinking tries to explain all their unusual and complex features by positive selection during adaptations to parasitism. Unfortunately, this straightforward approach does not withstand rigid scrutiny, which became possible only lately when molecular data from the sister non-parasitic lineages appeared. Indeed, even distantly related lineages of Alveolata and Kinetoplastea share several otherwise (very) rare molecular traits such as *trans*-splicing, polycistronic transcription and mitochondrial gene fragmentation and RNA editing [88,89]. The situation is even more complicated because many of these oddities apparently arose by convergent evolution through the accumulation of neutral mutations rather than by pure selection [90]. As exemplified by RNA editing, the molecular mechanisms underlying it are fundamentally different in Alveolata and Kinetoplastea, leading to the only plausible conclusion that these pathways have evolved in convergence [88]. In summary, the molecular data accumulated so far testify that many if not all of the peculiar features attributed to kinetoplastids were independently "tried on" at least several times during protistan evolution. Indeed, none of the traits appears to have evolved as a specific adaptation to the parasitic life style, as all of the typical trypanosomatid features are in some form already present in the basal kinetoplastid group of free-living bodonids.

However, considering the impressive success of trypanosomatid parasites, there might have been important pre-adaptations. One such example is the kDNA that is present in multiple fundamentally different forms in the bodonids [91], while a single kDNA network highly conserved in terms of organization, gene content and order, is a unifying feature of all trypanosomatids [92,93]. The transition from the non-catenated pro-kDNA of free-living *B. saltans* to the catenated kDNA disk of parasitic *P. confusum* might represent an important transition that contributed to the success of trypanosomatids, in which no further diversification of the kDNA seems to have occurred. In the absence of more information on the diverse bodonid kDNAs, we cannot pinpoint the critical differences, other than the potential importance of catenation of circular DNA molecules at the expense of their supercoiling [91]. A consequence of this invention might be a highly organized and efficient

replication and maintenance of a single kDNA catenane, allowing major contraction of the invariably huge kDNA of bodonids, which likely outweighed the inefficiency of replication by redundancy.

Trypanosomatids might serve as a good model group for tracing the evolution of parasitism. Several whole genomes of the dixenous trypanosomatids are already available for analysis [94–97]. By analyzing the genomic information from their free-living bodonid relatives and early-branching trypanosomatid *P. confusum* [28], we shall be able to identify features shared between these organisms as well as those representing key differences, especially in terms of gene loss and/or gain. From the currently available data, most differences are associated with genes encoding metabolic and cell surface proteins [43].

Modern methods of genome analysis allow direct comparison of the gene content between different groups of Trypanosomatida. The underlining assumption of this approach is that there must be some genes responsible for adaptation to the dixenous life style. Those genes should exclusively be either present or absent in *Trypanosoma*, *Leishmania*, and *Phytomonas* as compared to their monoxenous kins. Such an analysis promises to bring novel insight into what drives flagellates to the more complex dixenous life cycle and is indeed underway (Pavel Flegontov, J.L., V.Y., and J.V., in prep.).

Acknowledgements

This research was supported by the Czech Grant Agency P305/14/23986S, grant of the Czech Academy of Sciences M200961204, the Bioglobe grant CZ.1.07/2.3.00/30.0032, and the Praemium Academiae award to J. L., who is also a Fellow of the Canadian Institute for Advanced Research. V.Y. is supported by the Czech Grant Agency P506/13/24983S, the grant SGS25/PrF/2014 from the University of Ostrava, and funds of the Moravskoslezský Kraj research initiative, and T.S. received grant 108/2013/P from the University of South Bohemia. We acknowledge the use of research infrastructure that has received funding from the EU 7th Framework Programme under grant agreement No. 316304. We are thankful to Vladimír Hampl (Department of Parasitology, Charles University, Prague) and Marek Eliáš (Life Science Research Centre, University of Ostrava) for their help with Fig. 1.

References

[1] Pawlowski J, Audic S, Adl S, Bass D, Belbahri L, Berney C, et al. CBOL protist working group: barcoding eukaryotic richness beyond the animal, plant, and fungal kingdoms. PLoS Biol 2012;10:e1001419.

[2] Vickerman K. The evolutionary expansion of the trypanosomatid flagellates. Int J Parasitol 1994;24:1317–31.

[3] Adl SM, Simpson AG, Lane CE, Lukeš J, Bass D, Bowser SS, et al. The revised classification of eukaryotes. J Eukaryot Microbiol 2012;59:429–93.

[4] Moreira D, López-García P, Vickerman K. An updated view of kinetoplastid phylogeny using environmental sequences and a closer outgroup: proposal for a new classification of the class Kinetoplastea. Int J Syst Evol Microbiol 2004;54:1861–75.

[5] Simpson AG, Stevens JR, Lukeš J. The evolution and diversity of kinetoplastid flagellates. Trends Parasitol 2006;22:168–74.

[6] Tanifuji G, Kim E, Onodera NT, Gibeault R, Dlutek M, Cawthorn RJ, et al. Genomic characterization of *Neoparamoeba pemaquidensis* (Amoebozoa) and its kinetoplastid endosymbiont. Eukaryot Cell 2011;10:1143–6.

[7] Dyková I, Fiala I, Lom J, Lukeš J. *Perkinsiella* amoebae-like endosymbionts of *Neoparamoeba* spp., relatives of the kinetoplastid *Ichthyobodo*. Eur J Protistol 2003;39:37–52.

[8] von der Heyden S, Cavalier-Smith T. Culturing and environmental DNA sequencing uncover hidden kinetoplastid biodiversity and a major marine clade within ancestrally freshwater *Neobodo designis*. Int J Syst Evol Microbiol 2005;55:2605–21.

[9] Lukeš J, Arts GJ, van den Burg J, de Haan A, Opperdoes F, Sloof P, et al. Novel pattern of editing regions in mitochondrial transcripts of the cryptobiid *Trypanoplasma borreli*. EMBO J 1994;13:5086–98.

[10] Woo PT. *Cryptobia* (*Trypanoplasma*) *salmositica* and salmonid cryptobiosis. J Fish Dis 2003;26:627–46.

[11] Kumagai A, Ito H, Sasaki R. Detection of the kinetoplastid *Azumiobodo hoyamushi*, the causative agent of soft tunic syndrome, in wild ascidians *Halocynthia roretzi*. Dis Aquat Org 2013;106:267–71.

[12] Poynton SL, Whitaker BR, Heinrich AB. 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 2001;44:191–201.

[13] Hoare CA, Wallace FG. Developmental stages of trypanosomatid flagellates: a new terminology. Nature 1966;212:1385–6.

[14] Hoare CA. The classification of mammalian trypanosomes. Ergeb Mikrobiol Immunitatsforsch Exp Ther 1966;39:43–57.

[15] Wallace FG. The trypanosomatid parasites of insects and arachnids. Exp Parasitol 1966;18:124–93.

[16] McGhee RB, Cosgrove WB. Biology and physiology of the lower Trypanosomatidae. Microbiol Rev 1980;44:140–73.

[17] de Souza W. Structural organization of the cell surface of pathogenic protozoa. Micron 1995;26:405–30.

[18] de Souza W, Campanati L, Attias M. Strategies and results of field emission scanning electron microscopy (FE-SEM) in the study of parasitic protozoa. Micron 2008;39:77–87.

[19] Maslov DA, Votýpka J, Yurchenko V, Lukeš J. Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. Trends Parasitol 2013;29:43–52.

[20] Maslov DA, Lukeš J, Jirků M, Simpson L. Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. Mol Biochem Parasitol 1996;75:197–205.

[21] 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.

[22] Merzlyak E, Yurchenko V, Kolesnikov AA, Alexandrov K, Podlipaev SA, Maslov DA. Diversity and phylogeny of insect trypanosomatids based on small subunit rRNA genes: polyphyly of *Leptomonas* and *Blastocrithidia*. J Eukaryot Microbiol 2001;48:161–9.

[23] Westenberger SJ, Sturm NR, Yanega D, Podlipaev SA, Zeledon R, Campbell DA, et al. Trypanosomatid biodiversity in Costa Rica: genotyping of parasites from Heteroptera using the spliced leader RNA gene. Parasitology 2004;129:537–47.

[24] Votýpka J, Maslov DA, Yurchenko V, Jirků M, Kment P, Lun ZR, et al. Probing into the diversity of trypanosomatid flagellates parasitizing insect hosts in South-West China reveals both endemism and global dispersal. Mol Phylogenet Evol 2010;54:243–53.

[25] Votýpka J, Klepetková H, Jirků M, Kment P, Lukeš J. Phylogenetic relationships of trypanosomatids parasitising true bugs (Insecta: Heteroptera) in sub-Saharan Africa. Int J Parasitol 2012;42:489–500.

[26] Votýpka J, Klepetková H, Yurchenko VY, Horák A, Lukeš J, Maslov DA. Cosmopolitan distribution of a trypanosomatid *Leptomonas pyrrhocoris*. Protist 2012;163:616–31.

[27] Týč J, Votýpka J, Klepetková H, Šulaková H, Jirků M, Lukeš J. Growing diversity of trypanosomatid parasites of flies (Diptera: Brachcera): frequent cosmopolitism and moderate host specificity. Mol Phylogenet Evol 2013;69:255–64.

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

[29] Borghesan TC, Ferreira RC, Takata CS, Campaner M, Borda CC, Paiva F, et al. Molecular phylogenetic redefinition of *Herpetomonas* (Kinetoplastea, Trypanosomatidae), a genus of insect parasites associated with flies. Protist 2013;164:129–52.

[30] Jirků M, Yurchenko VY, Lukeš J, Maslov DA. New species of insect trypanosomatids from Costa Rica and the proposal for a new subfamily within the Trypanosomatidae. J Eukaryot Microbiol 2012;59:537–47.

[31] Yurchenko V, Votýpka J, Tesařová M, Klepetková H, Kraeva N, Jirků M, et al. Ultrastructure and molecular phylogeny of four new species of monoxenous trypanosomatids from flies (Diptera: Brachycera) with redefinition of the genus *Wallaceina*. Folia Parasitol 2014;61:97–112.

[32] Votýpka J, Suková E, Kraeva N, Ishemgulova A, Duží I, Lukeš J, et al. Diversity of trypanosomatids (Kinetoplastea: Trypanosomatidae) parasitizing fleas (Insecta: Siphonaptera) and description of a new genus *Blechomonas* gen. n. Protist 2013;164:763–81.

[33] von der Heyden S, Chao EE, Vickerman K, Cavalier-Smith T. Ribosomal RNA phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa. J Eukaryot Microbiol 2004;51:402–16.

[34] Finlay BJ, Fenchel T. Cosmopolitan metapopulations of free-living microbial eukaryotes. Protist 2004;155:237–44.

[35] Porcel BM, Denoeud F, Opperdoes FR, Noel B, Madoui M-A, Hammarton TC, et al. The streamlined genome of *Phytomonas* spp. relative to human pathogenic kinetoplastids reveals a parasite tailored for plants. PLoS Genet 2014;10:e1004007.

[36] Podlipaev SA. The more insect trypanosomatids under study-the more diverse Trypanosomatidae appears. Int J Parasitol 2001;31:648–52.

[37] Podlipaev SA. Insect trypanosomatids: the need to know more. Mem Inst Oswaldo Cruz 2000;95:517–22.

[38] Motta MC, Martins AC, de Souza SS, Catta-Preta CM, Silva R, Klein CC, et al. Predicting the proteins of *Angomonas deanei*, *Strigomonas culicis* and their respective endosymbionts reveals new aspects of the trypanosomatidae family. PLoS One 2013;8:e60209.

[39] Stevens JR. One million insects – a lot of parasites? Trends Parasitol 2001;17:119–20.
[40] Yurchenko V, Lukeš J, Jirků M, Maslov DA. Selective recovery of the cultivation-prone components from mixed trypanosomatid infections: a case of several novel species isolated from Neotropical Heteroptera. Int J Syst Evol Microbiol 2009;59:893–909.
[41] Yurchenko V, Lukeš J, Tesařová M, Jirků M, Maslov DA. Morphological discordance of the new trypanosomatid species phylogenetically associated with the genus *Crithidia*. Protist 2008;159:99–114.
[42] Lukeš J, Jirků M, Avliyakulov N, Benada O. Pankinetoplast DNA structure in a primitive bodonid flagellate, *Cryptobia helicis*. EMBO J 1998;17:838–46.
[43] Jackson AP, Quail MA, Berriman M. Insights into the genome sequence of a free-living Kinetoplastid: *Bodo saltans* (Kinetoplastida: Euglenozoa). BMC Genomics 2008;9:594.
[44] Maslov DA, Podlipaev SA, Lukeš J. Phylogeny of the kinetoplastida: taxonomic problems and insights into the evolution of parasitism. Mem Inst Oswaldo Cruz 2001;96:397–402.
[45] Stevens JR. Free-living bodonids and derived parasitic trypanosomatids: but what lies in between? Trends Parasitol 2014;30:113–4.
[46] Minchin EA. Investigations on the development of trypanosomes in tsetse flies and other Diptera. Q J Microsc Sci 1908;52:159–260.
[47] Poinar Jr G. Early Cretaceous trypanosomatids associated with fossil sand fly larvae in Burmese amber. Mem Inst Oswaldo Cruz 2007;102: 635–7.
[48] Poinar Jr G, Poinar R. Evidence of vector-borne disease of Early Cretaceous reptiles. Vector Borne Zoon Dis 2004;4:281–4.
[49] 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.
[50] Fernandes AP, Nelson K, Beverley SM. Evolution of nuclear ribosomal RNAs in kinetoplastid protozoa: perspectives on the age and origins of parasitism. Proc Natl Acad Sci USA 1993;90:11608–12.
[51] Noyes H. Implications of a Neotropical origin of the genus *Leishmania*. Mem Inst Oswaldo Cruz 1998;93:657–61.
[52] Noyes HA, Arana BA, Chance ML, Maingon R. The *Leishmania hertigi* (Kinetoplastida; Trypanosomatidae) complex and the lizard *Leishmania*: their classification and evidence for a neotropical origin of the *Leishmania-Endotrypanum* clade. J Eukaryot Microbiol 1997;44:511–7.
[53] Croan DG, Morrison DA, Ellis JT. Evolution of the genus *Leishmania* revealed by comparison of DNA and RNA polymerase gene sequences. Mol Biochem Parasitol 1997;89:149–59.
[54] Deane MP, Jansen AM. From a mono to a digenetic life-cycle: how was the jump for flagellates of the family Trypanosomatidae? Mem Inst Oswaldo Cruz 1988;83:273–5.
[55] Léger L. Sur les affinities de l'*Herpetomonas subulata* et la phylogenie des trypanosomes. Comp R Séances Soc Biol Ses Fil 1904;56:615–7.
[56] Chicharro C, Alvar J. Lower trypanosomatids in HIV/AIDS patients. Ann Trop Med Parasitol 2003;97(Suppl. 1):75–8.
[57] Ferreira MS, Borges AS. Some aspects of protozoan infections in immunocompromised patients – a review. Mem Inst Oswaldo Cruz 2002;97: 443–57.
[58] Rosenthal E, Marty P, del Giudice P, Pradier C, Ceppi C, Gastaut JA, et al. HIV and *Leishmania* coinfection: a review of 91 cases with focus on atypical locations of *Leishmania*. Clin Infect Dis 2000;31:1093–5.
[59] Dedet JP, Pratlong F. *Leishmania*, *Trypanosoma* and monoxenous trypanosomatids as emerging opportunistic agents. J Eukaryot Microbiol 2000;47:37–9.
[60] Pacheco RS, Marzochi MC, Pires MQ, Brito CM, Madeira Md, Barbosa-Santos EG. Parasite genotypically related to a monoxenous trypanosomatid of dog's flea causing opportunistic infection in an HIV positive patient. Mem Inst Oswaldo Cruz 1998;93:531–7.
[61] Sundar S, Chakravarty J. Recent advances in the diagnosis and treatment of kala-azar. Natl Med J India 2012;25:85–9.
[62] Bhattarai NR, Das ML, Rijal S, van der Auwera G, Picado A, Khanal B, et al. Natural infection of *Phlebotomus argentipes* with *Leishmania* and other trypanosomatids in a visceral leishmaniasis endemic region of Nepal. Trans R Soc Trop Med Hyg 2009;103:1087–92.
[63] Singh N, Chikara S, Sundar S. SOLiD sequencing of genomes of clinical isolates of *Leishmania donovani* from India confirm *Leptomonas* co-infection and raise some key questions. PLoS One 2013;8:e55738.
[64] Srivastava P, Prajapati VK, Vanaerschot M, Van der Auwera G, Dujardin JC, Sundar S. Detection of *Leptomonas* sp. parasites in clinical isolates of Kala-azar patients from India. Infect Genet Evol 2010;10:1145–50.
[65] Ghosh S, Banerjee P, Sarkar A, Datta S, Chatterjee M. Coinfection of *Leptomonas seymouri* and *Leishmania donovani* in Indian leishmaniasis. J Clin Microbiol 2012;50:2774–8.
[66] McGhee RB. The infection of avian embryos with *Crithidia* species and *Leishmania tarentola*. J Infect Dis 1959;105:18–25.
[67] De Sa MF, De Sa CM, Veronese MA, Filho SA, Gander ES. Morphologic and biochemical characterization of *Crithidia brasiliensis* sp. n. J Protozool 1980;27:253–7.
[68] Roitman I, Mundim MH, De Azevedo HP, Kitajima EW. Growth of *Crithidia* at high temperature: *Crithidia hutneri* sp. n. and *Crithidia luciliae thermophila* s. sp. n. J Protozool 1977;24:553–6.
[69] Nascimento MT, Garcia MC, da Silva KP, Pinto-da-Silva LH, Atella GC, Motta MC, et al. Interaction of the monoxenic trypanosomatid *Blastocrithidia culicis* with the *Aedes aegypti* salivary gland. Acta Trop 2010;113:269–78.
[70] Volf P, Hajmová M, Sádlová J, Votýpka J. Blocked stomodeal valve of the insect vector: similar mechanism of transmission in two trypanosomatid models. Int J Parasitol 2004;34:1221–7.
[71] Volf P, Hostomská J, Rohoušová I. Molecular crosstalks in *Leishmania*-sandfly-host relationships. Parasite 2008;15:237–43.
[72] Bates PA. Transmission of *Leishmania* metacyclic promastigotes by phlebotomine sand flies. Int J Parasitol 2007;37:1097–106.
[73] Molyneux DH. Vector relationships in the Trypanosomatidae. Adv Parasitol 1977;15:1–82.
[74] Siddall ME, Desser SS. Alternative leech vectors for frog and turtle trypanosomes. J Parasitol 1992;78:562–3.
[75] Soltys MA, Woo PT. Leeches as possible vectors for mammalian trypanosomes. Trans R Soc Trop Med Hyg 1968;62:154–6.
[76] 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.
[77] Austen JM, Ryan UM, Friend JA, Ditcham WG, Reid SA. Vector of *Trypanosoma copemani* identified as *Ixodes* sp. Parasitology 2011:1–7.
[78] Dougall AM, Alexander B, Holt DC, Harris T, Sultan AH, Bates PA, et al. Evidence incriminating midges (Diptera: Ceratopogonidae) as potential vectors of *Leishmania* in Australia. Int J Parasitol 2011;41:571–9.
[79] Kling JC, Korner H. Different regulatory mechanisms in protozoan parasitic infections. Int J Parasitol 2013;43:417–25.
[80] Haas M, Lukan M, Kišková J, Hrehová Z. Occurrence of blood parasites and intensity of infection in *Prunella modularis* in the montane and subalpine zone in the Slovak Carpathians. Acta Parasitol 2012;57:221–7.
[81] Zídková L, Čepička I, Szabová J, Svobodová M. Biodiversity of avian trypanosomes. Infect Genet Evol 2012;12:102–12.
[82] Hoare CA. Vampire bats as vectors and hosts of equine and bovine trypanosomes. Acta Trop 1965;22:204–16.
[83] Lizundia R, Newman C, Buesching CD, Ngugi D, Blake D, Sin YW, et al. Evidence for a role of the host-specific flea (*Paraceras melis*) in the transmission of *Trypanosoma* (*Megatrypanum*) pestanai to the European badger. PLoS One 2011;6:e16977.
[84] Holmes P. Tsetse-transmitted trypanosomes—their biology, disease impact and control. J Invertebr Pathol 2013;112(Suppl):S11–4.
[85] Rotureau B, Van Den Abbeele J. Through the dark continent: African trypanosome development in the tsetse fly. Front Cell Infect Microbiol 2013;3:53.
[86] Schnaufer A, Domingo GJ, Stuart K. Natural and induced dyskinetoplastic trypanosomatids: how to live without mitochondrial DNA. Int J Parasitol 2002;32:1071–84.
[87] Lai DH, Hashimi H, Lun ZR, Ayala FJ, Lukeš J. Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T. brucei*. Proc Natl Acad Sci USA 2008;105:1999–2004.
[88] Lukeš J, Leander BS, Keeling PJ. Cascades of convergent evolution: the corresponding evolutionary histories of euglenozoans and dinoflagellates. Proc Natl Acad Sci USA 2009;106(Suppl. 1):9963–70.
[89] Waller RF, Jackson CJ. Dinoflagellate mitochondrial genomes: stretching the rules of molecular biology. Bioessays 2009;31:237–45.
[90] Flegontov P, Gray MW, Burger G, Lukeš J. Gene fragmentation: a key to mitochondrial genome evolution in Euglenozoa? Curr Genet 2011;57:225–32.
[91] Lukeš J, Guilbride DL, Votýpka J, Zíková A, Benne R, Englund PT. Kinetoplast DNA network: evolution of an improbable structure. Euk Cell 2002;1:495–502.
[92] Liu B, Molina H, Kalume D, Pandey A, Griffith JD, Englund PT. Role of p38 in replication of Trypanosoma brucei kinetoplast DNA. Mol Cell Biol 2006;26:5382–93.
[93] Jensen RE, Englund PT. Network news: the replication of kinetoplast DNA. Annu Rev Microbiol 2012;66:473–91.
[94] Peacock CS, Seeger K, Harris D, Murphy L, Ruiz JC, Quail MA, et al. Comparative genomic analysis of three *Leishmania* species that cause diverse human disease. Nat Genet 2007;39:839–47.
[95] Ivens AC, Peacock CS, Worthey EA, Murphy L, Aggarwal G, Berriman M, et al. The genome of the kinetoplastid parasite, *Leishmania major*. Science 2005;309:436–42.
[96] Berriman M, Ghedin E, Hertz-Fowler C, Blandin G, Renauld H, Bartholomeu DC, et al. The genome of the African trypanosome *Trypanosoma brucei*. Science 2005;309:416–22.
[97] El-Sayed NM, Myler PJ, Blandin G, Berriman M, Crabtree J, Aggarwal G, et al. Comparative genomics of trypanosomatid parasitic protozoa. Science 2005;309:404–9.
[98] Kerr SF. Molecular trees of trypanosomes incongruent with fossil records of hosts. Mem Inst Oswaldo Cruz 2006;101:25–30.
[99] Lepage T, Bryant D, Philippe H, Lartillot N. A general comparison of relaxed molecular clock models. Mol Biol Evol 2007;24:2669–80.
[100] Yang Z, Rannala B. Bayesian estimation of species divergence times under a molecular clock using multiple fossil calibrations with soft bounds. Mol Biol Evol 2006;23:212–26.
[101] Gradstein FM, Ogg JG, Smith AG. A geologic time scale. Cambridge, UK: Cambridge University Press; 2004. p. 589.

5.3 PAPER III

Votýpka J, Rádrová J, **Skalický T**, Jirků M, Jirsová D, Mihalca AD, D'Amico G, Petrželková KJ, Modrý D, Lukeš J (2015) A tsetse and tabanid fly survey of African great apes habitats reveals the presence of a novel trypanosome lineage but the absence of *Trypanosoma brucei*. **International Journal of Parasitology** 45(12):741-8 (IF = 4.242).

Abstract

Tsetse and tabanid flies transmit several *Trypanosoma* species, some of which are human and livestock pathogens of major medical and socioeconomic impact in Africa. Recent advances in molecular techniques and phylogenetic analyses have revealed a growing diversity of previously unidentified tsetse-transmitted trypanosomes potentially pathogenic to livestock and/or other domestic animals as well as wildlife, including African great apes. To map the distribution, prevalence and co-occurrence of known and novel trypanosome species, we analyzed tsetse and tabanid flies collected in the primary forested part of the Dzanga-Sangha Protected Areas, Central African Republic, which hosts a broad spectrum of wildlife including primates and is virtually devoid of domestic animals. Altogether, 564 tsetse flies and 81 tabanid flies were individually screened for the presence of trypanosomes using 18S rRNA-specific nested PCR. Herein, we demonstrate that wildlife animals are parasitized by a surprisingly wide range of trypanosome species that in some cases may circulate via these insect vectors. While one-third of the examined tsetse flies harbored trypanosomes either from the *Trypanosoma theileri*, *Trypanosoma congolense* or *Trypanosoma simiae* complex, or one of the three new members of the genus *Trypanosoma* (strains ‘Bai’, ‘Ngbanda’ and ‘Didon’), more than half of the tabanid flies exclusively carried *T. theileri*. To establish the putative vertebrate hosts of the novel trypanosome species, we further analyzed the provenance of blood meals of tsetse flies. DNA individually isolated from 1033 specimens of *Glossina* spp. and subjected to high-throughput library-based screening proved that most of the examined tsetse flies engorged on wild ruminants (buffalo, sitatunga, bongo), humans and suids. Moreover, they also fed (albeit more rarely) on other vertebrates, thus providing indirect but convincing evidence that trypanosomes can be transmitted via these vectors among a wide range of warm- and cold-blooded hosts.

International Journal for Parasitology 45 (2015) 741–748

Contents lists available at ScienceDirect

International Journal for Parasitology

journal homepage: www.elsevier.com/locate/ijpara

A tsetse and tabanid fly survey of African great apes habitats reveals the presence of a novel trypanosome lineage but the absence of *Trypanosoma brucei* ☆

Jan Votýpka [a,b,*,1], Jana Rádrová [a,b,1], Tomáš Skalický [b,c,1], Milan Jirků [b,c], Dagmar Jirsová [c,d], Andrei D. Mihalca [e], Gianluca D'Amico [e], Klára J. Petrželková [b,f], David Modrý [b,d,g], Julius Lukeš [b,c,h,*]

[a] Department of Parasitology, Faculty of Science, Charles University, Prague, Czech Republic
[b] Institute of Parasitology, Biology Centre, České Budějovice (Budweis), Czech Republic
[c] Faculty of Sciences, University of South Bohemia, České Budějovice (Budweis), Czech Republic
[d] Central European Institute of Technology, University of Veterinary and Pharmaceutical Sciences, Brno, Czech Republic
[e] Department of Parasitology and Parasitic Diseases, Faculty of Veterinary Medicine, University of Agricultural Sciences and Veterinary Medicine, Cluj-Napoca, Romania
[f] Institute of Vertebrate Biology, Czech Academy of Sciences, Brno, Czech Republic
[g] Faculty of Veterinary Medicine, University of Veterinary and Pharmaceutical Sciences, Brno, Czech Republic
[h] Canadian Institute for Advanced Research, Toronto, Canada

ARTICLE INFO

Article history:
Received 5 February 2015
Received in revised form 17 June 2015
Accepted 22 June 2015
Available online 26 July 2015

Keywords:
Trypanosoma
Tsetse
Tabanids
African great apes
Gorillas
Transmission
Bloodmeal
Feeding preference

ABSTRACT

Tsetse and tabanid flies transmit several *Trypanosoma* species, some of which are human and livestock pathogens of major medical and socioeconomic impact in Africa. Recent advances in molecular techniques and phylogenetic analyses have revealed a growing diversity of previously unidentified tsetse-transmitted trypanosomes potentially pathogenic to livestock and/or other domestic animals as well as wildlife, including African great apes. To map the distribution, prevalence and co-occurrence of known and novel trypanosome species, we analyzed tsetse and tabanid flies collected in the primary forested part of the Dzanga-Sangha Protected Areas, Central African Republic, which hosts a broad spectrum of wildlife including primates and is virtually devoid of domestic animals. Altogether, 564 tsetse flies and 81 tabanid flies were individually screened for the presence of trypanosomes using 18S rRNA-specific nested PCR. Herein, we demonstrate that wildlife animals are parasitized by a surprisingly wide range of trypanosome species that in some cases may circulate via these insect vectors. While one-third of the examined tsetse flies harbored trypanosomes either from the *Trypanosoma theileri*, *Trypanosoma congolense* or *Trypanosoma simiae* complex, or one of the three new members of the genus *Trypanosoma* (strains 'Bai', 'Ngbanda' and 'Didon'), more than half of the tabanid flies exclusively carried *T. theileri*. To establish the putative vertebrate hosts of the novel trypanosome species, we further analyzed the provenance of blood meals of tsetse flies. DNA individually isolated from 1033 specimens of *Glossina* spp. and subjected to high-throughput library-based screening proved that most of the examined tsetse flies engorged on wild ruminants (buffalo, sitatunga, bongo), humans and suids. Moreover, they also fed (albeit more rarely) on other vertebrates, thus providing indirect but convincing evidence that trypanosomes can be transmitted via these vectors among a wide range of warm- and cold-blooded hosts.

© 2015 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

1. Introduction

The first scientific description of African trypanosomes was made at the end of the 19th century and the importance of the tsetse fly in transmission was discovered soon afterwards (Sharma et al., 2009). African trypanosomes are well-known causative agents of serious diseases in humans and livestock. A fatal human sleeping sickness occurs in sub-Saharan Africa and

☆ *Note:* Nucleotide sequence data reported in this paper are available in GenBank under accession numbers KP979582-4, KP979585-7, KR024686, KP307018-26, KR024687, KR024683-5, KR024688, SAMN03703425, SAMN03703430, SAMN03703434, SAMN03703438, SAMN03703509, SAMN03703516-8.

* Corresponding authors at: Department of Parasitology, Charles University, Viničná 7, 128 44 Prague, Czech Republic. Tel.: +420 221951826 (J. Votýpka). Biology Centre, Branišovská 31, 37005 České Budějovice, Czech Republic. Tel.: +420 387775416 (J. Lukeš).

E-mail addresses: vapid@natur.cuni.cz (J. Votýpka), jula@paru.cas.cz (J. Lukeš).

[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.ijpara.2015.06.005
0020-7519/© 2015 Australian Society for Parasitology Inc. Published by Elsevier Ltd. All rights reserved.

approximately 60 million people are exposed to the risk of infection, with more than 10,000 new cases reported annually (Simarro et al., 2011). Nagana, surra, and other animal trypanosomiases constitute such a serious problem for farming that they largely exclude productive breeds of livestock from parts of the humid and semi-humid zones of sub-Saharan African countries and annually cause approximately three million cattle deaths (Schofield and Kabayo, 2008).

The group of most significant African trypanosome species includes *Trypanosoma congolense*, *Trypanosoma vivax*, *Trypanosoma simiae* and five members of the *Trypanosoma brucei* complex. The diversity of African trypanosomes is much greater, however, and new species are continually described (Adams et al., 2006, 2010a; Hamilton et al., 2008). Nevertheless, life cycles, vertebrate hosts and vectors of many neglected and newly described trypanosome species remain unknown. It is clear that our knowledge of the diversity of African trypanosomes is limited and understanding their true diversity will require broader surveys.

Taxonomic designations of these diverse flagellates, mostly lacking species-specific morphological characters, are mostly based on their phylogenetic placements (Gibson, 2007). Moreover, trypanosome species may differ by only minor genetic differences, which reflect adaptations to particular parasitic niches. Recently, this has been exemplified by the case of *Trypanosoma evansi*, the life cycle of which is significantly influenced by just a few tiny genetic differences from *Trypanosoma brucei brucei* ancestral stock (Lai et al., 2008; Carnes et al., 2015).

Recent advances in molecular identification techniques and phylogenetic analyses have revealed previously unidentified trypanosome species in African vertebrates (Auty et al., 2012a) and tsetse flies (Adams et al., 2006, 2010a; Hamilton et al., 2008; Auty et al., 2012b). Due to the paucity of distinguishing morphological features, molecular tools comprise virtually the only way to map the high diversity of trypanosomes, yet only rarely are new forms formally described and named. Consequently, a growing number of nameless species represents an inconvenient complication, as these are largely ignored (Adams et al., 2010a).

While the mere presence of trypanosome DNA in bloodsucking flies cannot be considered proof of transmission by these vectors, it reveals their occurrence in the blood of vertebrates inhabiting the environment, in which collections occurred. New trypanosome species detected in tsetse flies could originate from wild vertebrates that are not yet known to host trypanosomes (Adams et al., 2010a) and, vice versa, undescribed trypanosomes found in wildlife could be transmitted by tsetse flies (Auty et al., 2012b). To track the entire life cycle and to reveal the bloodsucking vectors responsible for the transmission of these flagellates under natural conditions is time- and money-consuming and in some cases almost impossible. Nevertheless, the involvement of specific vertebrate hosts and invertebrate transmitters in the trypanosome life cycle could be uncovered by analyzing the blood meals and feeding patterns of possible vectors (Muturi et al., 2011). Not only do analyses of blood meals of tsetse flies trapped in their natural environments provide information about transmitted pathogens, but these hematophagous flies could also be used as "vertebrate samplers" and deliver DNA-based information on many aspects of vertebrate ecology (Calvignac-Spencer et al., 2013).

A great deal of African trypanosome biodiversity is probably still hiding in African savannahs and forests. Recent findings demonstrate that trypanosomes in and outside of the *T. brucei* complex can infect vertebrate hosts such as chimpanzees and other great apes that were previously believed to be more or less trypanosome-free (Jirků et al., 2015). To the best of our knowledge, however, no survey for trypanosomes in tsetse flies has been undertaken directly within habitats inhabited by African great apes. In this study, tsetse and tabanid flies captured in the Dzanga-Sangha Protected Areas (DSPA), Central African Republic (CAR), were examined to determine the diversity and prevalence of harbored trypanosomes and to survey the spectrum of vertebrates from which those flies took their blood meals.

2. Materials and methods

2.1. Study site

The DSPA in the CAR include zones with various levels of protection and restricted human access, as well as multiple-use zones with controlled human activities. The human population density in the DSPA is low (ca one person per km^2). The forested areas are composed of secondary and, less frequently, primary forest. Rainfall averages 1400 mm/year. Dry months typically occur between December and February, while the rest of the year has a long rainy season (albeit with low precipitation during June and July). Our sampling was carried out in the Dzanga Sector of the Dzanga-Ndoki National Park (2°50′N, 16°28′E) and its surroundings. Bai Hokou is a permanent Primate Habituation Programme research camp with local BaAka and Bantu trackers and research assistants, and up to five foreign volunteers and/or researchers. The surrounding area is covered by primary forest with several open spaces (forest glades), locally known as "bais". Two of these, Bai-Hokou (BH; 2°51′30.6″N, 16°28′12.6″E, in the proximity of the research camp) and Bai-Gubunga (BG; 2°50′59.9″N, 16°28′05.1″E, 3 km distant from the research camp), were chosen as the main collection sites of bloodsucking insects. Both locations are frequently visited by herds of forest buffalos (*Syncerus caffer nanus*), together with bongos (*Tragelaphus eurycerus*), sitatungas (*Tragelaphus spekii*) and forest elephants (*Loxodonta cyclotis*). The surrounding primary forest is inhabited by several duiker species (e.g., *Cephalophus silvicultor*), red river hogs (*Potamochoerus porcus*), giant forest hogs (*Hylochoerus meinertzhageni*), wild chimpanzees (*Pan troglodytes*), agile mangabeys (*Cercocebus agilis*) and several groups of western lowland gorillas (*Gorilla gorilla gorilla*) at different levels of habituation (Fig. 1).

2.2. Collection of insects

Tsetse and tabanid flies were collected in the DSPA during September 2012. The insects from the BH and BG collection sites were thoroughly examined by specific PCR for trypanosome prevalence and in the case of tsetse flies for blood source. The microhabitat of each collection and the occurrence of wildlife were noted. Most bloodsucking flies were collected on forest edges using tent-like Malaise traps with a black and dark-blue striped central wall and white roof that directs insects upwards to a cylinder containing a killing jar with 70% pure ethanol. The captured insects were collected from the Malaise traps twice each day. A small number of flies were obtained by sweep netting and hand collection. Ethanol-stored specimens were transported to the laboratory, where they were sorted under a stereomicroscope according to species, sex, collection site and date of trapping. Species identification was done using an online key (Les Glossines ou mouches tsé-tsé; http://www.cnev.fr/index.php/publications-et-outils/outils-didentification/identiciels) and confirmed by Dr. Pascal Grébaut (IRD-CIRAD, Montpellier, France). The correctness of the determination was further verified by barcoding five selected specimens from each morpho-species. This involved sequencing of their cytochrome *c* oxidase subunit 1 (*cox1*) gene using universal primers (Hebert et al., 2003), followed by a comparison with available databases, BOLD (http://www.boldsystems.org/) and GenBank (http://blast.ncbi.nlm.nih.gov/Blast.cgi).

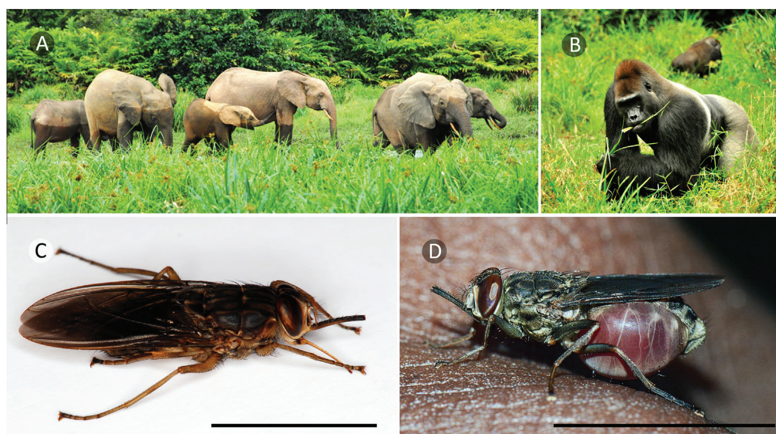


Fig. 1. Wild animals and tstse flies in Dzanga-Sangha Protected Areas, Central African Republic. (A) Herd of forest elephants in Bai-Gubunga. (B) Makumba, a silverback gorilla from a habituated group in Bai-Hokou. (C) *Glossina* (*Austenia*) *tabaniformis*. (D) *Glossina* (*Nemorhina*) *fuscipes fuscipes*. Scale bars = 1 cm.

2.3. DNA isolation, nested PCR, sequencing and recombinant sequence check

After washing individual tsetse and tabanid flies for 1 h in distilled water, their DNA was extracted using a High Pure PCR Template Preparation Kit (Roche, Germany) according to the manufacturer's instructions. Tissue homogenates were prepared from the whole tsetse or tabanid flies. In the case of tabanid flies, some specimens were divided into three body parts (head, thorax and abdomen), in order to determine the exact location of trypanosomes within the insect, and whether it is in accordance with the expected mode of transmission. For species identification, the trypanosome 18S rRNA genes were amplified using newly designed specific nested PCR with amplicon sizes reaching up to 2.2 kb and 2.1 kb in the first and second round, respectively. The first amplification round consisting of 35 cycles was performed in a final volume of 20 μl of PCR mix (TopBio, Czech Republic) containing 60–120 ng of total DNA isolated individually from the tsetse or tabanid fly body, and 10 pmol of each primer (S762: 5′-GACTTTTGCTTCCTCTAWTG-3′ and S763: 5′-CATATGCTTGTTTCA AGGAC-3′; Maslov et al., 1996). The annealing temperature was 55 °C for 90 s. For the next step consisting of 35 amplification cycles, 1 μl of amplified product was used as a template with 10 pmol of newly designed nested primers (TR-F2: 5′-GARTCTGC GCATGGCTCATTACATCAGA-3′ and TR-R2: 5′-CRCAGTTTGATGAGC TGCGCCT-3′). The annealing temperature was 64 °C for 90 s. Both strands of PCR products were sequenced. The search for possible recombinant sequences obtained by PCR was performed by Recombination Detection Program v.4 (RDP4) using the RDP, GENECONV, Chimaera, MaxChi, BOOTSCAN, SISCAN and 3seq methods with default settings (Martin et al., 2010).

2.4. Phylogenetic analysis

The 18S rRNA sequences obtained by nested PCR from flies were used to generate an alignment using K-align (http://www.ebi.ac.uk/Tools/msa/kalign/). Ambiguous positions and poorly alignable sequences were manually removed using BioEdit. The final 18S rRNA alignment included 2082 characters. An evolutionary model (TIM2+I+G) for this dataset was selected using the Akaike criterion in Modeltest 2.1.4 and used for Maximum Likelihood inference in PhyML 3.0. A heuristic search was performed using the subtree pruning and regrafting (SPR) branch-swapping algorithm and statistical support of bipartition was assessed using bootstrap resampling (1000 replicates) as described elsewhere (Votýpka et al., 2010, 2012). The accession numbers of sequences retrieved from GenBank and used in phylogenetic reconstructions and 18S rRNA alignments are provided in figures.

2.5. Preparation and analysis of the DNA library

To establish the origins of blood meals in tsetse flies, 1033 DNAs individually isolated from captured insects were separated into pools of 10 DNA samples and each pool was used as a template for specific PCR with primers Cytbb1F (5′-CCATCMAACATYTCADC ATGATGAAA-3′) and Cytbb2R (5′-GCHCCTCAGAATGAYATTTGKCCT CA-3′). PCR conditions were as follows: 94 °C for 5 min, then 35 cycles (94 °C for 1 min, 55 °C for 1 min, 72 °C for 1 min), and 72 °C for 7 min (Rádrová et al., 2013). Amplicons were purified using GenElute™ PCR Clean-Up (Sigma–Aldrich, Germany) and then further pooled according to tsetse fly species (*Glossina fuscipes fuscipes* versus *Glossina tabaniformis*), gender (male versus female), and capture site (BH versus BG) to form eight similarly numerous groups: (i) males of *G. f. fuscipes* from BH, (ii) females of *G. f. fuscipes* from BH, (iii) males of *G. tabaniformis* from BH, (iv) females of *G. tabaniformis* from BH, (v) males of *G. f. fuscipes* from BG, (vi) females of *G. f. fuscipes* from BG, (vii) males of *G. tabaniformis* from BG, and (viii) females of *G. tabaniformis* from BG.

DNA libraries from each group were prepared as described in the Rapid Library Preparation Method Manual (Roche) and then quantified using a TBS 380 Fluorometer (Turner, USA). Different Multiplex Identifiers (MIDs) with unique sequence tags were ligated to each group of amplicons and resulting DNA libraries were purified using Agencourt AMPure XP magnetic beads (Beckman Coulter, USA). Each DNA library was quantified using a TBS 380 Fluorometer (Turner); the number of molecules in each library was calculated and samples were diluted with Tris-EDTA buffer (10 mM TRIS, 1 mM EDTA, pH 8.0) to a concentration of 1×10^7 molecules/μl. Barcoded libraries were subsequently pooled and processed according to the emulsion PCR (emPCR)

Amplification Method Manual Lib-L (Roche). The library of DNA molecules was mixed with capture beads in a 1:1.7 ratio, emulsified and clonally amplified with emulsion-based PCR. Subsequently, beads carrying amplified DNA were recovered from emulsion, enriched and sequencing primers were annealed, yielding 450,000 beads. Sequencing using a GS Junior pyrosequencer (Roche) resulted in almost 47,000 raw sequences, which were analyzed and processed using the Mothur package (Schloss, 2009; www.mothur.org). Sequences were separated into eight groups according to their unique sequence tags present in MIDs. Denoising was performed by PyroNoise (minflows = 360; maxflows = 720; pdiffs = 3; bdiffs = 1) followed by sequence trimming (maxhomop = 8; minlength = 150). This resulted in 14,533 high quality sequences that were analyzed using BLASTN with an E-value cut-off set at 10-50. Statistical evaluation was performed using the Kruskal–Wallis and Mann–Whitney *U* tests (Statistica 6.1, StatSoft) and *P* values below 0.05 were considered significant. All primary sequencing data were deposited to the NCBI Sequence Read Archive under accession numbers SAMN03703425, SAMN03703430, SAMN03703434, SAMN03703438, SAMN03703509, SAMN03703516, SAMN03703517 and SAMN03703518. It should be noted that the methodology used is biased towards sequences from a recently taken blood meal where the host DNA is still highly abundant.

3. Results

3.1. Prevalence of trypanosomes and their 18S rRNA-based phylogeny

Morphological examination allowed allocation of all 1033 tsetse flies trapped in the two localities within the primary forest area of the DSPA to just two species. The subgenus *Austenia* was represented by 886 specimens of *G. tabaniformis*, while 147 specimens of *G. f. fuscipes* belonged to the subgenus *Nemorhina* (Fig. 1). All obtained *cox1* tsetse fly sequences clustered into two well-defined, internally homogenous groups (data not shown) fully corresponding with morphology-based determinations. Our *G. f. fuscipes* sequences (GenBank accession numbers KP979582-4) show 99% similarity with reference sequences. In the case of *G. tabaniformis* (KP979585-7), however, no sequences are publicly available and our sequences derived from this species constituted a separate clade among available tsetse fly sequences (data not shown).

The detection of trypanosomes by the nested PCR approach was performed on DNA samples isolated individually from a randomly selected subset of 564 tsetse flies representing 436 and 128 specimens of *G. tabaniformis* and *G. f. fuscipes*, respectively. The analyzed set was composed of 260 males and 202 females (102 specimens were homogenized prior to the determination of their sex), with 403 and 161 insects trapped at BH and BG, respectively. In total, 183 (32.4%) tsetse flies were positive for at least one trypanosome species (Table 1), with no significant correlation between infection and tsetse fly species ($\chi^2 = 0.684$, $P = 0.41$; 34.3% of *G. tabaniformis* versus 27.3% of *G. f. fuscipes*) or gender (*G. tabaniformis*: $\chi^2 = 2.073$, $P = 0.15$; 35.2% of males versus 28.1% of females; *G. f. fuscipes*: $\chi^2 = 0.406$, $P = 0.53$; 33.2% of males versus 27.7% of females) observed. The two trapping sites (BH and BG) differ in several aspects. First of all, they significantly differ in the distribution of the two tsetse fly species ($\chi^2 = 13.257$, $P < 0.0003$), with a relatively high occurrence of *G. f. fuscipes* in BH. Another notable difference is associated with species composition of vertebrates available in "bais" that serve as hosts for trypanosomes and as a source of blood for tsetse and tabanid flies. The BG locality is inhabited almost exclusively by herds of forest buffaloes, while other mammals visit both locations at approximately the same frequency (J. Votýpka, K.J. Petrželková, D. Modrý, personal observations).

The two sites differ significantly ($\chi^2 = 72.836$, $P < 0.0001$) in overall trypanosome prevalence, with 23.1% and 55.3% of tsetse flies from BH and BG, respectively, being positive. Regarding individual trypanosome species (see below), the number of trypanosome-positive tsetse flies was statistically significant only for *Trypanosoma theileri*. In congruence with overall trypanosome prevalence, tsetse flies were more commonly infected with *T. theileri* at BG (49.8%) than they were at BH (1.9%) ($\chi^2 = 91.187$, $P < 0.0001$) which is in good correlation with the occurrence of buffaloes in the former bai.

Altogether, based on 18S rRNA, we detected eight trypanosome species inventoried as shown in Table 1 and belonging to three subgenera (Fig. 2). This substantial diversity includes six members of the subgenus *Nannomonas*, one species affiliated with the subgenus *Trypanozoon*, and one species of the subgenus *Megatrypanum*. The tsetse fly species and prevalence of all registered trypanosomes are shown in Table 1. The more abundant *G. tabaniformis* hosted all eight trypanosome species, while the less numerous *G. f. fuscipes* was infected with only the three most prevalent trypanosomes (Table 1).

In our dataset, the subgenus *Nannomonas* is represented by three members of the *T. simiae* group: (i) the typical *T. simiae* (isolates G114 and G219 from *G. tabaniformis*; GenBank accession numbers KP307021 and KR024686, respectively); (ii) the new *T. simiae* 'Bai' (isolate G107 from *G. tabaniformis*; accession number KP307022), which forms a new subclade within the *T. simiae* complex together with the lineage *T. simiae* 'Tsavo'; and (iii) *Trypanosoma* sp. 'Fly9' (also called *T. godfreyi*-like in Adams et al., 2010a) (isolates G49 and G159 from *G. tabaniformis*; accession numbers KP307023 and KR024687, respectively). The latter trypanosome ('Fly9') was previously described from a tsetse fly (Malele et al., 2003) and is related to *Trypanosoma godfreyi* (Adams et al., 2010a). The species *T. congolense* is represented by the new lineage (iv) *T. congolense* 'Dzanga-Sangha' (isolates G22 from *G. tabaniformis*; accession number KP307024; and G34 and G39 from *G. f. fuscipes*; accession numbers KP307026 and KP307025, respectively). These three isolates form a new subclade within the species *T. congolense*, distinct from the hitherto recognized 'Forest', 'Savannah' and 'Kilifi' lineages. Finally, two closely related isolates within the subgenus *Nannomonas* that cluster between *T. congolense* and the *T. simiae* complex form two new species labeled (v) *Trypanosoma* sp. 'Ngbanda' (isolates G277 and G603 from *G. tabaniformis*; accession numbers KP307019 and KR024684, respectively) and (vi) *Trypanosoma* sp. 'Didon' (isolates G432 and G601 from *G. tabaniformis*; accession numbers KP307020 and KR024685, respectively).

The trypanosome labeled *Trypanosoma* sp. 'Makumba' (isolate G73 from *G. f. fuscipes* and isolate G433 from *G. tabaniformis*; accession numbers KP307018 and KR024683, respectively) is closely related to *Trypanosoma* sp. 'Msubugwe' (Hamilton et al., 2008), with which it constitutes a sister group of the subgenus *Trypanozoon* (Fig. 2). Finally, the subgenus *Megatrypanum* is represented by the *T. theileri* complex (isolate G24 from *G. f. fuscipes*; accession number KR024688), although as the sequences obtained are slightly heterogeneous, these may represent several genotypes or a complex of closely related species (data not shown). In addition to tsetse flies, we captured 81 female tabanids belonging to the genera *Chrysops* (43 specimens), *Haematopota* (6), and *Tabanus* (32), of which 48 (59.3%) were positive for *T. theileri* (Table 2). No other trypanosome species was detected. Prevalence of infection, which was confined to the thoracic and abdominal parts (Fig. 3; also see Section 4), differed significantly among genera ($\chi^2 = 10.502$, $P < 0.001$; 44.2% and 81.3% positivity for *Chrysops* and *Tabanus*, respectively). While 46.8% of tabanids

Table 1
Prevalence of trypanosomes in tsetse flies captured in the Dzanga-Sangha Protected Areas, Central African Republic.

Tsetse fly species	Examined/positive (%)	Prevalence of *Trypanosoma* spp. (%)							
		T.s.	*T.s.*B	Fly9	*T.c.*	Ngb.	Did.	Mak.	*T.t.*
Glossina tabaniformis	436/149 (34.2)	2.8	0.7	0.7	8.7	0.5	0.5	5.0	15.1
Glossina fuscipes fuscipes	128/35 (27.3)	0.0	0.0	0.0	7.8	0.0	0.0	2.3	17.2
Total	564/183 (32.4)	2.1	0.5	0.5	8.5	0.4	0.4	4.4	15.6

T.s., typical *Trypanosoma simiae*; *T.s.*B, *Trypanosoma simiae* 'Bai'; Fly9, *Trypanosoma* sp. 'Fly9'; *T.c.*, *Trypanosoma congolense* 'Dzanga-Sangha'; Ngb., *Trypanosoma* sp. 'Ngbanda'; Did., *Trypanosoma* sp. 'Didon'; Mak., *Trypanosoma* sp. 'Makumba'; *T.t.*, *Trypanosoma theileri*.

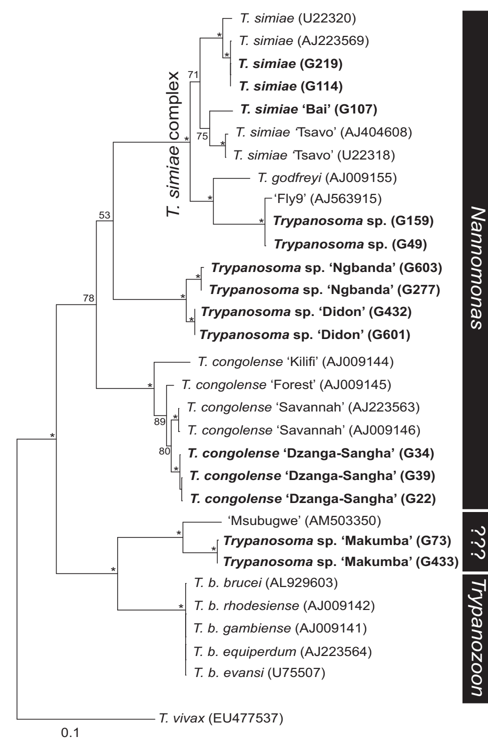


Fig. 2. 18S rRNA-based phylogeny of trypanosomes from tsetse flies. Maximum likelihood tree with bootstrap values for 1000 replicates (*indicates value >95). 18S rRNA sequences generated within this study are in bold. Each sequence, labeled by tsetse fly number (e.g., G219), belongs to a given species/complex/group (see Table 1). Notes on the newly described species: *Trypanosoma simiae* 'Bai' is a new member of the *T. simiae* complex; *Trypanosoma congolense* 'Dzanga-Sangha' is a new lineage of the species *T. congolense*; *Trypanosoma* sp. 'Ngbanda' and *Trypanosoma* sp. 'Didon' are two new species constituting a new clade within the subgenus *Nannomonas*; and *Trypanosoma* sp. 'Makumba' is a new member of the *Trypanosoma* sp. 'Msubugwe' group. Sequences obtained belonging to the *Trypanosoma theileri* complex of the subgenus *Megatrypanum* were not included.

were positive for trypanosomes at BH, positivity increased to 81.5% at BG ($\chi^2 = 7.717$, $P < 0.005$).

In order to identify possible recombinant sequences generated by PCR, we performed recombination detection analysis with the RDP4 program using the RDP, GENECONV, Chimaera, MaxChi, BOOTSCAN, SISCAN and 3seq detection algorithms. Application of these algorithms failed to identify any statistically significant recombination events ($P < 0.05$).

3.2. Determination of tsetse fly blood meal

To establish upon which vertebrates the tsetse flies fed and whether this varied by collection site, tsetse fly species and/or sex, the set of 1,033 DNAs individually isolated from specimens captured at BH and BG was subdivided into eight categories. Next, an equal amount of DNA from each category was pooled and used for the preparation of eight separate Lib-L libraries (see subsection 2.5). Due to substantial differences in the number of sequences extracted from each library, only limited mutual comparison was possible. Furthermore, it should be noted that the data obtained do not represent the blood "prevalence" or the number of tsetse flies engorging on given hosts, but only the frequencies of different sequences that could to some extent be affected by such factors as blood freshness. The strength of the method used lies in its ability to detect rare sequences. Indeed, it was able to monitor a wide range of blood sources.

Eight species of warm- and two species of cold-blooded vertebrates constituted more than 99% of the 14,533 high-quality sequences obtained from the engorged tsetse flies (Table 3). The dominant source of blood at both sampled sites consisted of forest buffaloes (*Syncerus caffer nanus*; 42.9%), with humans (24.9%) representing the second most frequent source (Fig. 1). At BG, a location distant from any human settlements, human blood occurred significantly less frequently (16.0%). Three groups of large forest mammals represented an additional important source of blood: bongo and sitatunga (*Tragelaphus* spp.; 15.7%), two species of wild hogs (*Potamochoerus porcus* and *Hylochoerus meinertzhageni*; 12.7%), and duikers (*Cephalophus* spp.; 4.1%), whereas the blood of rodents (unspecified member of the order Rodentia), forest elephants (*Loxodonta cyclotis*), tortoises (Testudinidae), and crocodiles (*Osteolaemus tetraspis*) was identified only sporadically (Table 3). The sporadic presence of elephant blood is noteworthy, as these are frequent visitors to both examined sites (Fig. 1).

4. Discussion

Recently, infections by trypanosomes mainly belonging to the *T. brucei* complex were detected with unexpected frequency in tissue samples and feces of African great apes (Jirků et al., 2015). To evaluate the possibility that tsetse and/or tabanid flies take

Table 2
Prevalence of trypanosomes in tabanid flies captured in the Dzanga-Sangha Protected Areas, Central African Republic.

Tabanid genus	Examined/positive (%)
Chrysops	43/19 (44.2)
Haematopota	6/3 (50.0)
Tabanus	32/26 (81.3)
Total	81/48 (59.3)

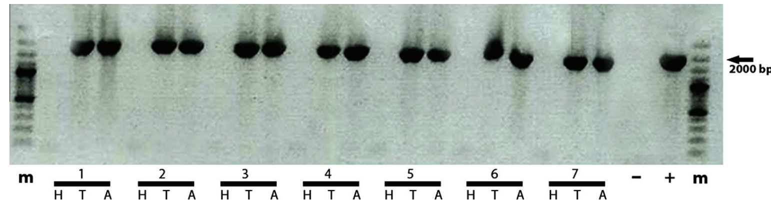


Fig. 3. Detection of *Trypanosoma theileri* in body parts of *Tabanus* spp. Each individual was dissected into three body parts: head (H), thorax (T) and abdomen (A). Trypanosomes were detected by PCR only in the thorax and abdomen. 1–7, tabanid specimen number; m, marker; –, negative and +, positive controls.

Table 3
Host cytochrome *b* gene detected in tsetse flies captured in the Dzanga-Sangha Protected Areas, Central African Republic.

Host	Collection site		Tsetse fly species		Tsetse fly sex		Seq. analysis
	BH (No: 582)	BG (No: 451)	G.t. (No: 886)	G.f.f. (No: 147)	Male (No: 367)	Female (No: 556)	No. of seq. (%)
African buffalo	40.66%[a]	68.91%[a]	50.62%	50.15%	53.26%	50.62%	2364 (42.88)
Human	24.87%[a]	16.04%[a]	23.19%	26.12%	24.03%	23.86%	1685 (30.56)
Bongo/Sitatunga	15.77%	11.46%	12.28%	19.57%	17.24%	16.38%	875 (15.87)
Bush pig/Giant hog	12.69%[a]	1.92%[a]	9.15%[a]	1.04%[a]	2.71%	3.68%	364 (6.60)
Duiker	4.40%	1.07%	3.41%	2.88%	1.75%	4.69%	173 (3.14)
Unidentified rodent	0.99%	0.42%	0.91%	0.14%	1.01%	0.26%	26 (0.47)
Forest elephant	0.30%	N.D.	0.18%	N.D.	N.D.	0.22%	10 (0.18)
Chicken	0.07%	0.05%	0.05%	0.05%	N.D.	0.05%	4 (0.07)
Unidentified tortoise	0.07%	N.D.	0.05%	N.D.	N.D.	0.00%	2 (0.04)
Dwarf crocodile	N.D.	0.03%	0.03%	N.D.	N.D.	0.03%	1 (0.02)
Total number of seq.	2495	2091	3306	2207	1763	2826	5513 (100)

BH, Bai-Hokou; BG, Bai-Gubunga; G.t., *Glossina tabaniformis*; G.f.f., *Glossina fuscipes fuscipes*; No, number of analyzed tsetse flies; seq., sequences; N.D., not detected.
[a] Statistically significant difference ($P < 0.05$).

blood from apes and are involved in transmitting their trypanosomes, an extensive analysis of these potential vectors from a habitat of chimpanzees and gorillas was performed.

Upon searching the literature, we came to realize that tsetse and tabanid flies had never before been collected in a primary forest inhabited by African great apes and other wildlife. For that reason, we undertook an extensive search for trypanosomes in hundreds of flies. Employing high-throughput screening using next-generation sequencing of almost all tsetse flies enabled us to determine the spectrum of animal species upon which the flies had fed. Indeed, at both studied sites within the DSPA, the DNA of several mammalian species was detected in *G.* (*Austenia*) *tabaniformis* and *G.* (*Nemorhina*) *fuscipes fuscipes*, thereby proving the opportunistic feeding behavior of *Glossina* flies (Muturi et al., 2011). Feeding of *G. fuscipes fuscipes* and several other tsetse species on cold-blooded vertebrates such as monitor lizard, turtle and snake has been documented previously (Waiswa et al., 2006; Bouyer et al., 2007; Farikou et al., 2010) and is further corroborated by our data.

Although lowland gorillas and tsetse flies co-habit in both capture sites and the Malaise traps were applied in the bais at least twice during visits by these great apes, they do not appear to be frequent targets of the examined *Glossina* spp.. Such an observation is non-trivial and surprising, especially when the virtual absence of blood from gorillas and other apes is compared with the abundant presence of human blood. Indeed, both *Glossina* spp. were often seen to attack humans tracking the gorillas (J. Votýpka, K.J. Petrželková, D. Modrý, personal observations; Fig. 1). One possible explanation for the obtained results implies the existence of some unknown anti-tsetse fly behavior by gorillas and chimpanzees, however, other explanations must also be considered. Host odor is an important component of tsetse fly perception and these insects do not seem to be attracted nor repelled by the primate smell. Another possibility is that the relative abundance of great apes in the sampling area is still low compared with other available hosts, and even analysis of 1000 tsetse flies is insufficient to cover the whole host spectrum and exhaustively map feeding preferences of the studied tsetse flies. Finally, because the effectiveness of black and dark-blue striped traps of various shapes differs significantly among tsetse fly species, a confounding presence of other, non-attracted and consequently undetected, *Glossina* spp. cannot be excluded.

Recently, several new trypanosome species or genotypes have been discovered in tsetse flies (Adams et al., 2006, 2010b; Hamilton et al., 2008) and large mammals (Auty et al., 2012a), gradually increasing the diversity and host range of African trypanosomes. Regrettably, these unnamed flagellates remain largely ignored, probably due to an absence of formal taxonomic description (Adams et al., 2010a). The reasons for this fact include unavailability in culture and the lack of identified vertebrate and/or insect hosts, which may well include rare species or those that have been inaccessible to investigation.

Our nested 18S rRNA-based PCR assay proved to be highly sensitive and provided sufficient information for a phylogenetic analysis that revealed an unexpectedly broad spectrum composed of at least eight trypanosome operational taxonomic units (OTUs). However, the presence of trypanosome DNA in the blood meals of tsetse flies is not proof of their capacity to transmit the parasite. Some OTUs are very closely related to species of medical and veterinary importance. At least one of these trypanosomes was detected in almost one-third of the examined tsetse flies, testifying to its active circulation in the DSPA. Two representatives of the subgenus *Nannomonas*, *T. simiae* and *T. congolense*, recently expanded into a complex of four and three lineages, respectively (Njiru et al., 2004; Hamilton et al., 2008; Adams et al., 2010a,b; Simo et al., 2012). Our findings further increase this diversity:

Trypanosoma simiae 'Bai' extends the *T. simiae* complex into five lineages (typical *T. simiae*, *T. simiae* 'Tsavo', *T. simiae* 'Bai', *T. godfreyi* and *Trypanosoma* sp. 'Fly9') and the herein newly described *T. congolense* 'Dzanga-Sangha' expands the species *T. congolense* into four lineages (*T. congolense* 'Forest', 'Savannah', 'Kilifi' and 'Dzanga-Sangha'). Finally, *Trypanosoma* sp. 'Makumba' clusters together with *Trypanosoma* sp. 'Msubugwe' (Hamilton et al., 2008; Adams et al., 2010a,b) to form an enigmatic group which may represent an evolutionary link between the subgenera *Trypanozoon* and *Nannomonas*.

Two additional novel flagellates of the subgenus *Nannomonas* first reported in this study, *Trypanosoma* sp. 'Ngbanda' and *Trypanosoma* sp. 'Didon' form, in the 18S rRNA-based tree, a new clade branching off between the *T. congolense* and *T. simiae* complexes. Such a position of this new clade raised concern about a potential hybrid molecule. Due to a very low amount of DNA, we failed to amplify additional genes (e.g. glyceraldehyde 3-phosphate dehydrogenase). However, we have obtained the same results twice and independently from different tsetse fly specimen, and further analyzed the amplified sequences using the recombinant detection RDP4 programs, which failed to identify any statistically significant recombination events. Therefore, we believe the obtained sequences are authentic.

We recently established a diagnostic ITS1-based PCR assay that allows detection of even very small amounts of trypanosome DNA in ethanol- or RNAlater-stored feces or in partially decomposed tissue samples collected from dead animals (Jirků et al., 2015). While useful for assignment to previously known species, the information obtained in the ITS1 region is insufficient for phylogenetic analyses of novel lineages which should, if possible, be based on the 18S rRNA gene.

Using the ITS1-based PCR assay, the *T. brucei* complex was frequently detected in an extensive survey of African great apes from several tropical African countries but not CAR (Jirků et al., 2015). The 18S rRNA-based nested PCR assay revealed circulation of at least eight trypanosome lineages, but we did not detect any trypanosomes belonging to the *T. brucei* complex. This result can be explained either (i) by an overall absence of this complex in the studied biotope (no information is available about *T. brucei* in humans and animals in the studied area), or (ii) by the presence of an unregistered tsetse fly species transmitting flagellates of the *T. brucei* complex.

No significant differences were observed in trypanosome prevalence among tsetse fly species or gender, although significant differences did exist between the two trapping sites in terms of overall trypanosome prevalence and, in particular, the abundance of the most common species, *T. theileri*. Since the same trend was observed in the trapped tabanid flies, we believe that these differences correspond with the higher concentration at Bai-Gubunga (BG) of forest buffaloes, of which *T. theileri* is a frequent parasite.

The extremely high positivity of tabanid flies for the *T. theileri* is likely a consequence of their increased life span compared with other insect vectors, resulting in a high number of blood meals. Furthermore, the striking differences in the prevalence of *T. theileri* can be explained by the large and highly aggressive tabanid flies of the genus *Tabanus* preferentially attacking forest buffaloes, while the relatively small members of the genus *Chrysops* seems to prefer smaller forest ungulates. The invariable restriction of *T. theileri* to the gut (Fig. 3) of these important vectors is fully compatible with the stercorarian-type development proposed by Bose and Heister (1993).

In summary, the present study confirms the feeding opportunism of tsetse flies. Indeed, an extensive analysis of their blood meal not only identified their main mammalian targets but also confirmed previous evidence for their capacity to feed on cold-blooded vertebrates. The absence of gorillas, which occur frequently at both studied locations, in the feeding range of those tsetse flies analyzed may be due to the relatively low number of analyzed flies, the presence of unregistered tsetse fly species, or some anti-tsetse fly behavior by the apes. Finally, we have demonstrated that tsetse flies, but not tabanid flies, harbor a surprisingly broad spectrum of trypanosomes, some of which were encountered, to our knowledge, for the first time.

Acknowledgements

Research permission was obtained from the relevant ministry of the Central African Republic. Special thanks are given to A. Todd, the park staff, the Primate Habituation Programme and the World Wildlife Fund of the Dzanga-Sangha Protected Areas, Central African Republic. We would also like to thank J. Košnar and A. Chro ňáková from the Institute of Soil Biology, Czech Republic, for their help with Next Generation sequencing. This study was supported by grant number M200961204 and the Praemium Academiae award from the Academy of Sciences of the Czech Republic, grant 108/2013/P from the University of South Bohemia, Czech Republic, and European Union 7th Framework Programme grants CZ.1.07/2. 3.00/300032 and 316304. This research was also co-financed by the European Social Fund and state budget of the Czech Republic (CZ.1.07/2.3. 00/20.0300), UEFISCDI grant PCE236/2011 and COST Action TD1303.

References

Adams, E.R., Hamilton, P.B., Gibson, W.C., 2010a. African trypanosomes: celebrating diversity. Trends Parasitol. 26, 324–328.

Adams, E.R., Hamilton, P.B., Rodrigues, A.C., Malele, I.I., Delespaux, V., Teixeira, M.M.G., Gibson, W., 2010b. New *Trypanosoma* (*Duttonella*) *vivax* genotypes from tsetse flies in East Africa. Parasitology 137, 641–650.

Adams, E.R., Malele, I.I., Msangi, A.R., Gibson, W.C., 2006. Trypanosome identification in wild tsetse populations in Tanzania using generic primers to amplify the ribosomal RNA ITS-1 region. Acta Tropica 100, 103–109.

Auty, H., Anderson, N.E., Picozzi, K., Lembo, T., Mubanga, J., Hoare, R., Fyumagwa, R.D., Mable, B., Hamill, L., Cleaveland, S., Welburn, S.C., 2012a. Trypanosome diversity in wildlife species from the Serengeti and Luangwa Valley ecosystems. PLoS Negl. Trop. Dis. 6 (10), e1828. http://dx.doi.org/10.1371/journal.pntd.0001828.

Auty, H.K., Picozzi, K., Malele, I., Torr, S.J., Cleaveland, S., Welburn, S.C., 2012b. Using molecular data for epidemiological inference. assessing the prevalence of *Trypanosoma brucei rhodesiense* in tsetse in Serengeti, Tanzania. PLoS Negl. Trop. Dis. 6 (1), e1501. http://dx.doi.org/10.1371/journal.pntd.0001501.

Bose, R., Heister, N.C., 1993. Development of *Trypanosoma* (*M.*) *theileri* in tabanids. J. Euk. Microbiol. 40, 788–792.

Bouyer, J., Pruvot, M., Bengaly, Z., Guerin, P.M., Lancelot, R., 2007. Learning influences host choice in tsetse. Biol. Lett. 3, 113–116.

Calvignac-Spencer, S., Leendertz, F.H., Gilbert, M.T., Schubert, G., 2013. An invertebrate stomach's view on vertebrate ecology: certain invertebrates could be used as "vertebrate samplers" and deliver DNA-based information on many aspects of vertebrate ecology. Bioessays 35, 1004–1013. http://dx.doi.org/10.1002/bies.201300060.

Carnes, J., Anupama, A., Balmer, O., Jackson, A., Lewis, M., Brown, R., Cestari, I., Desquesnes, M., Gendrin, C., Hertz-Fowler, C., Imamura, H., Ivens, A., Kořený, L., Lai, D.-H., MacLeod, A., McDermott, S.M., Merritt, C., Monnerat, S., Moon, W., Myler, P.J., Phan, I., Ramasamy, G., Sivam, D., Lun, Z.-R., Lukeš, J., Stuart, K., Schnaufer, A., 2015. Genome and phylogenetic analyses of *Trypanosoma evansi* reveal extensive similarity to *T. brucei* and multiple independent origins for dyskinetoplasty. PLoS Negl. Trop. Dis. 9, e3404. http://dx.doi.org/10.1371/journal.pntd.0003404.

Farikou, O., Njiokou, F., Simo, G., Asonganyi, T., Cuny, G., Geiger, A., 2010. Tsetse fly blood meal modification and trypanosome identification in two sleeping sickness foci in the forest of southern Cameroon. Acta Trop. 116, 81–88.

Gibson, W., 2007. Resolution of the species problem in African trypanosomes. Int. J. Parasitol. 37, 829–838.

Hamilton, P.B., Adams, E.R., Malele, I.I., Gibson, W.C., 2008. A novel, high-throughput technique for species identification reveals a new species of tsetse-transmitted trypanosome related to the *Trypanosoma brucei* subgenus, *Trypanozoon*. Infect. Genet. Evol. 8, 26–33.

Hebert, P.D.N., Cywinska, A., Ball, S.L., DeWaard, J.R., 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B Biol. Sci. 270, 313–321.

Jirků, M., Votýpka, J., Petrželová, K.J., Jirků-Pomajbíková, K., Kriegová, E., Vodička, R., Junglen, S., Lankaster, F., Leendertz, S.A.J., Wittig, R.M., Boesch, C., Modrý, D., Ayala, F.J., Leendertz, F.H., Lukeš, J., 2015. Wild chimpanzees are infected by *Trypanosoma brucei*. Int. J. Parasitol. Parasit. Wildl. 4, 277–282.

Lai, D.H., Hashimi, H., Lun, Z.R., Ayala, F.J., Lukeš, J., 2008. Adaptations of *Trypanosoma brucei* to gradual loss of kinetoplast DNA: *Trypanosoma equiperdum* and *Trypanosoma evansi* are petite mutants of *T. brucei*. Proc. Natl. Acad. Sci. U.S.A. 105, 1999–2004.

Malele, I., Craske, L., Knight, C., Ferris, V., Njiru, Z.K., Hamilton, P., Lehane, S., Lehane, M., Gibson, W., 2003. The use of specific and generic primers to identify trypanosome infections of wild tsetse flies in Tanzania by PCR. Infect. Genet. Evol. 3, 271–279.

Martin, D.P., Lemey, P., Lott, M., Moulton, V., Posada, D., Lefeuvre, P., 2010. RDP3: a flexible and fast computer program for analyzing recombination. Bioinformatics 26, 2462–2463.

Maslov, D.A., Lukeš, J., Jirků, M., Simpson, L., 1996. Phylogeny of trypanosomes as inferred from the small and large subunit rRNAs: implications for the evolution of parasitism in the trypanosomatid protozoa. Mol. Biochem. Parasitol. 75, 197–205.

Muturi, C.N., Ouma, J.O., Malele, I.I., Ngure, R.M., Rutto, J.J., Mithofer, K.M., Enyaru, J., Masiga, D.K., 2011. Tracking the feeding patterns of tsetse flies (*Glossina* genus) by analysis of bloodmeals using mitochondrial cytochromes genes. PLoS One 6, e17284. http://dx.doi.org/10.1371/journal.pone.0017284.

Njiru, Z.K., Makumi, J.N., Okoth, S., Ndungu, J.M., Gibson, W.C., 2004. Identification of trypanosomes in *Glossina pallidipes* and *G. longipennis* in Kenya. Infect. Genet. Evol. 4, 29–35.

Rádrová, J., Šeblová, V., Votýpka, J., 2013. Feeding behavior and spatial distribution of *Culex* mosquitoes (Diptera: Culicidae) in wetland areas of the Czech Republic. J. Med. Entomol. 50, 1097–1104.

Simarro, P.P., Diarra, A., Postigo, J.A.R., Franco, J.R., Jannin, J.G., 2011. The human African trypanosomiasis control and surveillance programme of the World Health Organization 2000–2009: the way forward. PLoS Negl. Trop. Dis. 5, e1007. http://dx.doi.org/10.1371/journal.pntd.0001007.

Simo, G., Silatsa, B., Flobert, N., Lutumba, P., Mansinsa, P., Madinga, J., Manzambi, E., De Deken, R., Asonganyi, T., 2012. Identification of different trypanosome species in the mid-guts of tsetse flies of the Malanga (Kimpese) sleeping sickness focus of the Democratic Republic of Congo. Parasites Vectors 5, 201. http://dx.doi.org/10.1186/1756-3305-5-201.

Schloss, P.D., 2009. A high-throughput DNA sequence aligner for microbial ecology studies. PLoS One 4, e8230. http://dx.doi.org/10.1371/journal.pone.0008230.

Schofield, C.J., Kabayo, J.P., 2008. Trypanosomiasis vector control in Africa and Latin America. Parasites Vectors 1, 24. http://dx.doi.org/10.1186/1756-3305-1-24.

Sharma, R., Gluenz, E., Peacock, L., Gibson, W., Gull, K., Carrington, M., 2009. The heart of darkness: growth and form of *Trypanosoma brucei* in the tsetse fly. Trends Parasitol. 25, 517–524.

Votýpka, J., Maslov, D.A., Yurchenko, V., Jirků, M., Kment, P., Lun, Z.R., Lukeš, J., 2010. Probing into the diversity of trypanosomatid flagellates parasitizing insect hosts in South-West China reveals both endemism and global dispersal. Mol. Phylogenet. Evol. 54, 243–253.

Votýpka, J., Szabová, J., Rádrová, J., Zídková, L., Svobodová, M., 2012. *Trypanosoma culicavium* sp. nov., an avian trypanosome transmitted by *Culex* mosquitoes. Int. J. Syst. Evol. Microbiol. 62, 745–754.

Waiswa, C., Picozzi, K., Katunguka-Rwakishaya, E., Olaho-Mukani, W., Musoke, R.A., Welburn, S.C., 2006. *Glossina fuscipes fuscipes* in the trypanosomiasis endemic areas of south eastern Uganda: Apparent density, trypanosome infection rates and host feeding preference. Acta Trop. 99, 23–29.

5.4 MANUSCRIPT I

Skalický T, Dobáková E, Wheeler R, Tesařová M, Flegontov P, Jirsová D, Votýpka J, Yurchenko V, Ayala FJ, Lukeš J (submitted manuscript) Extensive flagellar remodeling during the complex life cycle of *Paratrypanosoma*, an early-branching trypanosomatid.

Abstract

Paratrypanosoma confusum is a monoxenous kinetoplastid flagellate that constitutes the most basal branch of the highly diverse parasitic trypanosomatids, which include human pathogens *Trypanosoma* and *Leishmania*. This makes *Paratrypanosoma* uniquely informative for the evolution of obligatory parasitism from free-living life style and the evolution of human parasitism in some trypanosomatid lineages. It has typical promastigote morphology, but also forms surface-attached haptomonads and amastigotes. The haptomonad stage forms by attachment to a surface *via* a large bulge at the base of the flagellum, which is then remodeled into a thin attachment pad associated with flagellum shortening. Promastigotes and haptomonads multiply by binary division, and the progeny of a haptomonad can either remain attached or grow a flagellum and resume swimming. Whole genome sequencing and transcriptome profiling, in combination with analysis of the cell ultrastructure, reveal how the cell surface and metabolism are adapted to parasitism, and how characteristic cytoskeletal features are conserved. Our data demonstrate that surface attachment by the flagellum and the flagellar pocket, a *Leishmania*-like flagellum attachment zone and a *T. cruzi*-like cytostome are ancestral features, while evolution of extant kinetoplastids, including the human parasites, is associated with genome streamlining and diversification of membrane proteins.

Extensive flagellar remodeling during the complex life cycle of *Paratrypanosoma*, an early-branching trypanosomatid

Tomáš Skalický [1,2,*], Eva Dobáková [1,&,*], Richard Wheeler [3,*], Martina Tesařová [1], Pavel Flegontov [1,4], Dagmar Jirsová [1,$], Jan Votýpka [1,5], Vyacheslav Yurchenko [1,4,6], Francisco J. Ayala [7,#] & Julius Lukeš [1,2,8,#]

[1] Institute of Parasitology, Biology Centre, Czech Academy of Sciences and [2] Faculty of Science, University of South Bohemia, 37005 České Budějovice, Czech Republic; [3] Sir William Dunn School of Pathology, University of Oxford, Oxford, OX1 3RE, U. K.; [4]Faculty of Science, University of Ostrava, 71000 Ostrava, Czech Republic; [5] Department of Parasitology, Faculty of Science, Charles University, 12844 Prague, Czech Republic; [6] Faculty of Science, University of Ostrava, 71000 Ostrava, Czech Republic; [7] Department of Ecology and Evolutionary Biology, University of California, Irvine, CA, 92697, USA; [8] Canadian Institute for Advanced Research, Toronto, ON M5G1Z8, Canada [*] Equal contribution [#] Corresponding authors (jula@paru.cas.cz and fjayala@uci.edu) [&] Present address: Faculty of Natural Sciences, Comenius University, Bratislava, Slovakia [$] Present address: Faculty of Science, Mendel University, Brno, Czech Republic

Submitted to Proceedings of the National Academy of Sciences of the United States of America

***Paratrypanosoma confusum* is a monoxenous kinetoplastid flagellate that constitutes the most basal branch of the highly diverse parasitic trypanosomatids, which include human pathogens *Trypanosoma* and *Leishmania*. This makes *Paratrypanosoma* uniquely informative for the evolution of obligatory parasitism from free-living life style and the evolution of human parasitism in some trypanosomatid lineages. It has typical promastigote morphology, but also forms surface-attached haptomonads and amastigotes. The haptomonad stage forms by attachment to a surface *via* a large bulge at the base of the flagellum, which is then remodeled into a thin attachment pad associated with flagellum shortening. Promastigotes and haptomonads multiply by binary division, and the progeny of a haptomonad can either remain attached or grow a flagellum and resume swimming. Whole genome sequencing and transcriptome profiling, in combination with analysis of the cell ultrastructure, reveal how the cell surface and metabolism are adapted to parasitism, and how characteristic cytoskeletal features are conserved. Our data demonstrate that surface attachment by the flagellum and the flagellar pocket, a *Leishmania*-like flagellum attachment zone and a *T. cruzi*-like cytostome are ancestral features, while evolution of extant kinetoplastids, including the human parasites, is associated with genome streamlining and diversification of membrane proteins. Author contribution: J.L. and F.J.A. designed research; T.S., E.D., M.T., D.J., J.V. and V.Y. performed research; T.S., R.W., P.F., and J.L. analyzed data; T.S., R.W., P.F., F.J.A. and J.L. wrote the paper.**

trypanosomatid | evolution | cytostome | flagellar remodeling | haptomonads

INTRODUCTION

Kinetoplastid flagellates are diverse and widespread protists, best known for serious human diseases caused by *Trypanosoma* and *Leishmania*. Indeed, members of these two and other genera belong to the most successful parasites, with a unique and wide range of adaptations to the host environment. It was proposed that the disease-causing trypanosomatids with the two-host (= dixenous) life style, which combines the insect vector and the mammalian or plant host, evolved from flagellates parasitizing solely insects (1, 2). The earliest known branch of the trypanosomatid clade, predating its massive diversification, is *Paratrypanosoma confusum* found in mosquites (3), and a free-living clade closest to trypanosomatids is the genus *Bodo* (4, 5).

The biflagellated, bacterivorous *Bodo*, from which parasitic trypanosomatids emerged, has almost twice as many genes as the monoflagellated parasitic trypanosomatids (5). To identify further features associated with the evolution of parasitism, we analyzed the morphology of *Paratrypanosoma* and its adaptation to different *in vitro* environments which, combined with analysis of the genome and transcriptome, allowed identification of genes possibly associated with these features. The single flagellum of trypanosomatids is a highly flexible structure used for locomotion, attachment and sensing. Its structure is also subject to substantial restructuring during the life cycle to adapt to different functions (6-8). Flagellar motility is required for transmission via a vector, immune evasion (9) and cell division (10) of *Trypanosoma brucei* and is intimately associated with the vital flagellar pocket structure in all trypanosomatids. Recently, additional functions of this dexterous cellular component have been described, such as production of the extracellular vesicles which may mediate host interaction (11) and parasite-parasite interaction by membrane exchange or fusion (12). In the juxtaform morphological superclass that includes trypomastigotes and epimastigotes, the flagellum is laterally attached to the cell by an extended flagellum attachment zone (FAZ). Alternatively, the flagellum may protrude from the flagellar pocket without an extended attachment in the liberform morphological superclass that includes promastigotes, opisthomastigotes and choanomastigotes (13, 14).

Here we report that *Paratrypanosoma* and the stercorarian trypanosomes, such as *Trypanosoma cruzi* and *T. grayi*, retain more

Significance

Kinetoplastids are a group of protists with unique morphology and molecular features. Many of them developed parasitic lifestyle and are economically and medically important causative agents of serious crop, animal and human diseases. Interspersed between parasitic trypanosomatids and free-living bodonids, *Paratrypanosoma confusum* is uniquely informative for study of the emergence of parasitism. It is morphologically very flexible, as it forms three distinct life stages that can be studied separately. Particularly interesting is the haptomonad stage which rebuilds its flagellum into an extensive adhesive plaque. As an adaptation to parasitism, *Paratrypanosoma* lost a plethora of enzymes involved in breakdown of macromolecules and the capacity of receptor-mediated endocytosis, but has gained surface proteins and membrane transporters to obtain nutrients from the host.

Reserved for Publication Footnotes

Submission PDF

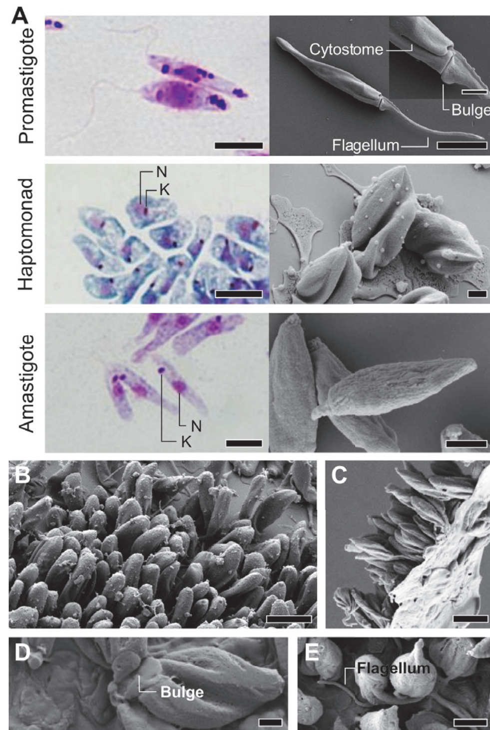


Fig. 1. Morphologies of *Paratrypanosoma*. A) Different morphologies by light microscopy of Giemsa-stained cells (left) and SEM (right). Promastigote has a long flagellum with a prominent bulge at its base (inset). The kinetoplast (K) and nucleus (N) are indicated. Neither haptomonad nor amastigote have a long flagellum. **B-E)** SEM of haptomonads. **B)** Haptomonads attach perpendicular to the surface in dense clusters. **C)** Haptomonad cluster detached from the surface by propylene oxide showing the underside of the attachment pad. **D)** Haptomonads retain a bulge at the tip of their short flagellum. **E)** Occasional long flagella are visible extending from the haptomonad attachment pad. Scale bars: 5 µm (A, B, C; except in A middle image - 10 µm, and inset -1 µm); D) 1 µm; E) 2 µm.

ancestral genes than the other trypanosomatid clades investigated. Despite lacking lateral attachment of the flagellum to the cell body as in *Trypanosoma*, *Paratrypanosoma* has a complex cell-cell body attachment *via* a FAZ similar to *Leishmania*. It also has a complex cytostome/flagellar pocket architecture similar to that of *T. cruzi* but evidently lost in *T. brucei* and *L. mexicana*. Furthermore, it has the capacity to restructure the flagellum to attach to surfaces, generating an extensive flagellum-derived adhesive pad, and does so readily in culture allowing the first analysis of this process, which occurs in most trypanosomatid lineages (15-17). Judging by the distribution of morphological traits in trypanosomatids, we suppose that *Paratrypanosoma* morphology is close to that of the last common ancestor of trypanosomatids.

RESULTS

The discovery of *Paratrypanosoma* as an outgroup to dixenous trypanosomes and all other trypanosomatids calls for a study focused on the distribution of morphological and molecular traits

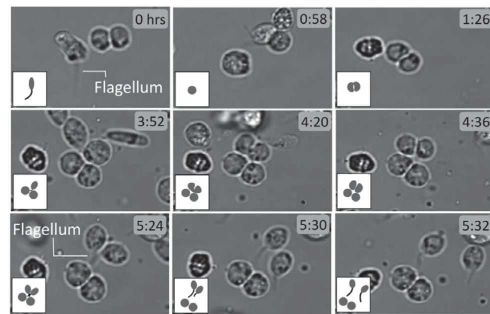


Fig. 2. Promastigote attachment and replication as a haptomonad. Still images from time-lapse videos of promastigote and haptomonad interconversion. Attachment of a promastigote to form a haptomonad and two successive divisions; first forming two haptomonads, then two haptomonads and two promastigotes. Time (hours:minutes) for each frame is shown in the top right, and a cartoon of cell arrangement is shown in the bottom left.

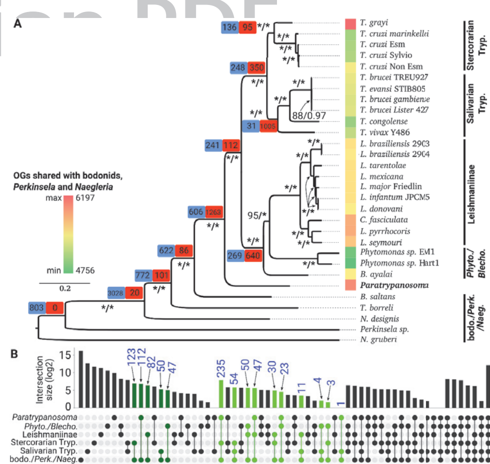


Fig. 3. Ancestral gene families in *Paratrypanosoma* and other trypanosomatids. A) Phylogenetic tree based on a concatenated set of 98 proteins, constructed with the maximum likelihood and Bayesian approaches (see Methods for details). For each species, the counts of orthologous groups shared with at least one bodonid and/or *Naegleria* are color-coded (see a scale on the left). Support values for tree nodes are shown in the format 'bootstrap support / posterior probability', and full support is indicated with asterisks. Gains (in blue) and losses (in red) of OGs were mapped on the tree using the Dollo parsimony algorithm. Few basal nodes are shown for clarity (see a full tree with gains/losses mapped in Suppl. Fig. 6). **B)** Phyletic patterns visualized for 6 clades: *Paratrypanosoma*, *Phytomonas/Blechomonas*, Leishmaniinae (*Leishmania*, *Leptomonas*, *Crithidia*), stercorarian trypanosomes, salivarian trypanosomes, and bodonids/*Perkinsela*/*Naegleria*. Counts of OGs unique for each clade and all possible intersections of the 6 sets are shown in the bar plot using a log scale (see selected OG counts above the bars). Counts of ancestral gene families shared by one or two trypanosomatid clades are highlighted in dark-green and light-green, respectively.

on the trypanosomatid tree. Light microscopy of live and Giemsa-stained cells and scanning electron microscopy (SEM) showed that in liquid medium axenic *Paratrypanosoma* attains both a motile promastigote with a long free flagellum (Fig. 1A) and

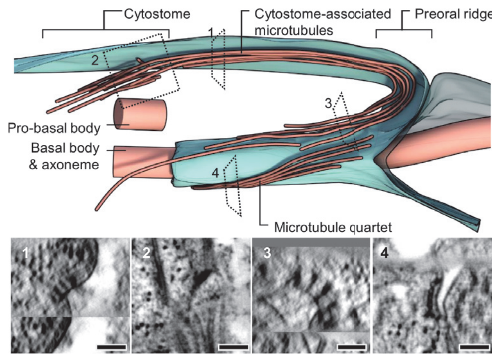


Fig. 4. Paratrypanosoma has a cytostome similar to T. cruzi and a pocket architecture similar to Leishmania. Three-dimensional model of the pocket/cytostome of Paratrypanosoma assembled from an electron tomogram. Virtual sections through the preoral ridge (1), the cytostome (2), the cytostome microtubules exiting the pocket (3) and the FAZ-associated microtubule quartet (4) are shown. Scale bars indicate 100 nm.

Submission PDF

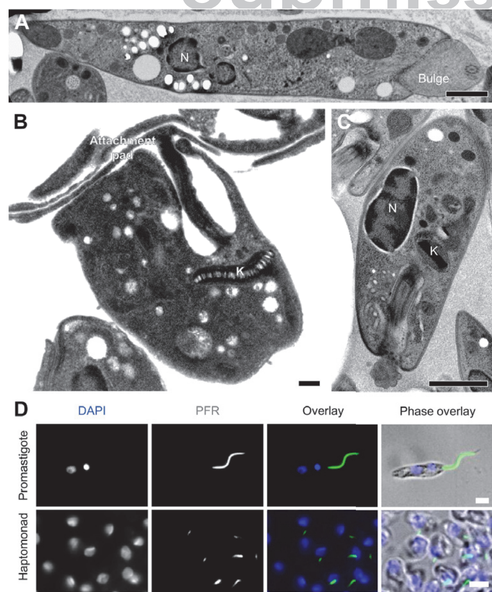


Fig. 5. Restructuring of the *Paratrypanosoma* flagellum between promastigotes, haptomonads and amastigotes. A-C) TEM showing the morphology of the flagellum in **A)** promastigote, **B)** haptomonad and **C)** amastigote. The attachment pad in haptomonad is continuous with the flagellum. **D)** Immunofluorescence of promastigotes and haptomonads using an anti-PFR antibody. The PFR is reduced or absent in haptomonads. Scale bars indicate 1 μm (A and C); 250 nm (B) and 2 μm (D).

a sedentary, surface attached stage somewhat similar to haptomonads of *Leishmania*. The haptomonad stage forms extensive thin attachment plaques on the plastic substrate (Figs. 1B-E). We have observed these morphologies also in the gut of the

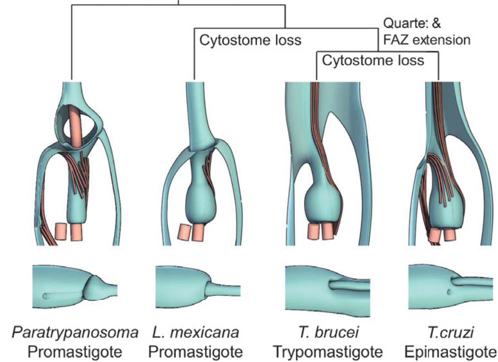


Fig. 6. Evolution of the flagellar pocket/cytostome complex of human infective trypanosomatids. Cartoon summarizing the likely loss of the cytostome and extension of the microtubule quartet and FAZ to generate *Leishmania*, *T. brucei* and *T. cruzi* pocket/cytostome morphology from an ancestral *Paratrypanosoma*-like morphology.

Culex quinquefasciatus host. On semi-solid agar, *Paratrypanosoma* attains a third morphology: an amastigote without a noticeable free flagellum (Fig. 1A). These morphologies are typical of monoxenous trypanosomatids, with widespread incidences of surface attachment in the gut of the invertebrate host and of amastigotes when excreted. They lack FAZ to the side of the cell body, confirming *Paratrypanosoma* is a part of the liberform morphological superclass (13).

We have been unable to ascertain the relation of axenic culture cell morphologies to the *Paratrypanosoma* life cycle, which is not well understood. In *C. quinquefasciatus* mosquitoes infected either by feeding on sugar or blood meal, *Paratrypanosoma* was detected in the gut and/or crop on the 1st (100% specimens, n = 8), 2nd (100%, n = 9) and 3rd day post-infection (90%, n = 10), but not after defecation. Gaining and retention of the infection from 3rd or 4st instar larvae through pupation to adult mosquitos was tested by addition of promastigotes into water, where they survive for at least 2 days. The emerged mosquitoes of both sexes were never infected (n = 80). This suggests defective infectivity of the culture-adapted parasites. Possible dixenous life cycle and transmission through a vertebrate host was tested on 4 BALB/c mice by a combination of subcutaneous and intraperitoneal injection of 10^7 promastigotes, however, 1 to 4 weeks post-infection parasites were detected neither by cultivation nor PCR. Further support for a monoxenous life style comes from the failure of *Paratrypanosoma* to survive at 37°C for more than a few hours. Nonetheless, life cycles with promastigote, haptomonad and amastigote stages are common among insect-parasitizing trypanosomatids and arise in *Leishmania* vectors, making analysis of *Paratrypanosoma* morphology and ultrastructure *in vitro* valuable.

Surface attachment is widespread among trypanosomatids; yet the development of the plaque used for attachment (Figs. 1B-E) was never studied in detail. Therefore, we analyzed the alternation of *Paratrypanosoma* between the promastigote and haptomonad morphologies in culture by time-lapse light microscopy (Fig. 2) and compared them to the morphology of haptomonads by SEM (Figs. 1B-E). Surface attachment of promastigotes typically occurred by the base of the external flagellum, the 'bulge' (Figs. 1A, D). Following attachment, the flagellum then shortened to a negligible rod barely extending from the pocket, on the time scale of 1 h (Fig. 2; Suppl. Fig. 1A). This was associated with a repositioning of the cell to an upright orientation, and a co-

incident spreading of the bulge into a thin extended pad attached to the surface. A colony of haptomonads could cover a large area by these extended pads. Time-lapse light microscopy indicated haptomonads can divide while attached, generating daughter cells that may remain attached or grow an external flagellum and detach (Fig. 2; Suppl. Fig. 1B; Suppl. Video 1 and 2). Cell division takes approximately 6 h at 20°C, while elongation of the flagellum and detachment occurs on the time scale of 1 to 2 h (Fig. 2).

Transformation from promastigotes into haptomonads was promoted by nutrient availability and alkaline pH. The digestive tract of a mosquito is nutrient-rich and typically alkaline (18), and is thus suitable for haptomonads. The attachment pad itself was highly resistant to all our attempts to degrade it enzymatically (SI Datasets S1), indicating it could confer strong attachment even under harsh digestive conditions of the mosquito gut. These morphological transitions appear sensitive to nutrients in the environment and, like life cycle transitions in other species, include significant remodeling of the flagellum. We therefore sequenced the *Paratrypanosoma* genome as a reference for transcriptome analysis of different morphological stages, for insight into metabolism, and as a comparative resource, to analyze evolution of flagellum and flagellum-related cytoskeleton.

Using paired-end and mate-pair Illumina reads, we assembled a draft genome with 31× average coverage, 31.4 Mbp, 2,114 scaffolds, maximum scaffold length 2.99 Mbp, and scaffold N50 of 438 Kbp. Using Augustus trained on a set of unambiguous gene models for conserved genes and relying also on RNA-seq read mapping, we annotated 8,668 protein-coding genes, 66 tRNAs and 122 copies of rRNA genes. This gene complement is similar to that of other trypanosomatids (19). A substantial fraction of core eukaryotic genes (72.3%) was found in *Paratrypanosoma* (compare with high-quality genome assemblies: 74.9% in *T. brucei*, 73.6% in *L. major* and 72.6% in *L. pyrrhocoris*), indicating that the assembly is essentially complete (20).

Comparing gene expression between haptomonads and promastigotes revealed 327 and 264 genes significantly upregulated in haptomonads and promastigotes, respectively (false discovery rate corrected p-value < 0.05, fold change > 2) (SI Datasets S2). Functional annotation of these genes revealed that products of the most upregulated genes in haptomonads are associated with the ribosome, its biogenesis or translation (SI Datasets S3A), while those upregulated in promastigotes are part of intermediary metabolism or redox processes (SI Datasets S3B).

Next, we turned to comparative genomics to investigate whether *Paratrypanosoma* has retained more ancestral features than other trypanosomatids. Using the OrthoFinder (21), we defined orthologous groups (OGs) of proteins for a large set of trypanosomatids, for bodonids (free-living *Bodo saltans* and *Neobodo designis*, a parasite *Trypanoplasma borreli*, and an endosymbiont *Perkinsela* sp.) and for *Naegleria gruberi*, a heterolobosean (Fig. 3; SI Datasets S4). Next, we counted the number of ancestral OGs in each species, i.e. those shared with any bodonid or *Naegleria*. *Trypanosoma grayi*, a parasite of crocodiles that belongs to the clade of stercorarian trypanosomes, shows the highest number of ancestral OGs, 6,197, while *Paratrypanosoma* shows the second highest number, 6,066. Given that both species have the shortest branches in a multigene tree (Fig. 3), this result is not unexpected. Thus, we have reasons to believe these two species are the slowest-evolving trypanosomatids in the analyzed dataset.

We also scrutinized the distribution of ancestral OGs across major mono- or paraphyletic groups: i/ *Paratrypanosoma*, ii/ salivarian and iii/ stercorarian trypanosomes, iv/ Leishmaniinae (*Leishmania*, *Leptomonas*, *Crithidia*), v/ *Phytomonas*/*Blechomonas* and vi/ bodonids/*Perkinsela*/*Naegleria*. Stercorarian trypanosomes and *Paratrypanosoma* have the highest numbers of ancestral OGs unique to those clades, 123 and 112, respectively. When ancestral OGs occurring in any two trypanosomatid clades are considered, *Paratrypanosoma* and stercorarian trypanosomes co-inherit by far the highest number of such groups (235), while other intersections contain 54 groups or less (Fig. 3). Although most *Paratrypanosoma* genes having the two phyletic patterns described above are annotated as hypothetical proteins or lack specific annotations (Suppl. Tables S5 and S6), we made the following observations: 1/ some genes of tryptophan and histidine catabolism, and arginine biosynthesis are unique to *Paratrypanosoma* and/or the stercorarian trypanosomes and were lost in the other trypanosomatids; 2/ proteins of the dispersed gene family 1 (DFG1) (22) are unique to *Paratrypanosoma* and the stercorarian trypanosomes. Remarkably, being the fifth largest protein family in *T. cruzi*, these are long membrane proteins of unknown function stored in intracellular vesicles, with their extracellular domains excreted during the trypomastigote to amastigote transformation (22, 25). The DFG1 proteins are upregulated at the *T. cruzi* amastigote stage as compared to the trypomastigote and epimastigote stages (22), while in *Paratrypanosoma*, they tend to be upregulated at the haptomonad stage (SI Datasets S2).

Finally, we looked at the relationship of phyletic patterns and functional categories for OGs. Using a one-way ANOVA analysis combined with Tukey's honest significance test, we show that differential gene expression at the haptomonad and promastigote stages depends on phyletic patterns of corresponding OGs (p-value = $1.22×10^{-7}$; SI Datasets S7). The ancestral genes co-inherited by the stercorarian trypanosomes and *Paratrypanosoma* and lost in the other clades tend to be upregulated in haptomonads: there are 84 genes in this group with statistically significant changes in their expression, and among them 55 are upregulated in haptomonads, and only 29 in promastigotes (SI Datasets S2). Stage-specific expression of this group is significantly different from that of the universally conserved genes, genes shared by *Paratrypanosoma* and both *Trypanosoma* clades only (to the exclusion of bodonids), and *Paratrypanosoma*-specific genes (p-values adjusted for multiple testing = 0.012, 0.006, $8×10^{-7}$, respectively). Genes within two latter groups tend to be upregulated in promastigotes. In general, ancestral genes with various phyletic patterns are either constitutively expressed in *Paratrypanosoma* or upregulated in haptomonads. On the contrary, trypanosomatid-specific genes tend to be upregulated at the promastigote stage (SI Datasets S7). These results suggest that the haptomonad stage might be an ancestral character in trypanosomatids.

The dramatic morphological change between promastigotes and haptomonads is exemplified by the bulge at the flagellum base for flagellum attachment, which in haptomonads expands to form an extensive adhesive plaque. Since this may involve restructuring of the paraflagellar rod (PFR), FAZ and flagellar axoneme, we analyzed conservation of known proteins, primarily identified in *T. brucei*, forming these structures across kinetoplastids (Suppl. Fig. 2). The FAZ proteins are of particular interest as they appear to have adaptable functions in cell morphogenesis. They were first identified in the extended FAZ of *Trypanosoma* but emerged as abundant components of the flagellar pocket neck in *Leishmania* promastigotes.

Using TEM and electron tomography we analyzed the structure of the pocket and base of the flagellum to determine the ultrastructural features responsible for the haptomonad morphological adaptation. TEM revealed extensive attachment of the bulge to the cell body by desmosome-like structures in promastigotes and haptomonads, comparable to the *Leishmania* flagellar pocket neck, albeit covering a larger area (Suppl. Figs. 3A, B). This suggests the proteins involved in bulge-cell attachment are the FAZ proteins, and RNA-seq indeed showed FAZ mRNAs were present in both promastigotes and haptomonads. Attach-

Submission PDF

ment is particularly complex in the distal pocket region, likely mediated by FAZ10, and this attachment was elaborated in haptomonads (Suppl. Fig. 3B). Immunofluorescence using the anti-*T. brucei* FAZ antibody Dot1 identified a structure in promastigotes near the expected localization of FAZ10, which also showed elaboration in haptomonads (Suppl. Fig. 3C). The *Paratrypanosoma* genome encodes orthologs of almost all FAZ proteins (Suppl. Fig. 2A), while *Bodo* and *Neobodo* have orthologs of around half of them. The trypanosomatids therefore appear to have diversified FAZ proteins. Given the ancestral trypanosomatid was likely liberform, as judged from the distribution of the liberform and juxtaform characters in this group, we propose that the FAZ proteins originally evolved to generate the haptomonad morphotype. The extended FAZ of the juxtaform *Trypanosoma* later arose in that lineage. Some FAZ proteins are often lost among liberforms (FAZ4, FAZ13), and some OGs (FLA, FLABP, FAZ11) show duplication among juxtaforms. Further candidates for forming the extended FAZ may be identified among OGs gained at the *Trypanosoma* node (Suppl. Tables S5, S8).

SEM of promastigotes revealed a cytostome-like indentation near the cell anterior (Fig. 1A; Suppl. Fig. 4A). TEM of the pocket structure (Suppl. Fig. 4B) and 3D reconstruction of the pocket organization by electron tomography showed that overall it was typical of promastigotes, including those of *Leishmania*, with a simple invagination surrounded by complex electron-dense areas and sets of microtubules (Fig. 4). There were two sets of specialized microtubules; a quartet similar to the FAZ quartet of *T. brucei* and *Leishmania*, and a highly-decorated set of microtubules associated with the cytostome. These ran from the pocket neck around the preoral ridge, back to a dip in the cell surface from which microtubules extend into the cytoplasm (Fig. 4; Suppl. Fig. 4A). This structure is comparable to the *T. cruzi* cytostome, suggesting the *Leishmania*-like flagellar pocket neck/FAZ structure and the microtubule quartet were ancestral and later expanded. It also indicates the cytostome was present in the ancestral trypanosomatid, has been lost in *Leishmania*, many monoxenous trypanosomatids and salivarian trypanosome lineages, but retained in the stercorarian trypanosomes and some monoxenous parasites, including *Crithidia fasciculata*. No gains or losses of FAZ OGs suggested a function in formation of the cytostome.

The *Paratrypanosoma* promastigote flagellum has the canonical 9+2 axoneme and, based on the genome sequence, the canonical molecular composition, too (Suppl. Fig. 2). Both haptomonads and amastigotes have greatly shortened flagella. In *Leishmania*, flagellar shortening during the promastigote/amastigote transition is associated with additional axoneme remodeling, loss of the central pair, distal motor proteins and radial spokes giving a transition from a 9+0 to a collapsed 9v (variable) axoneme. We used TEM to check whether similar axoneme restructuring occurred in the *Paratrypanosoma* haptomonad or amastigote (Fig. 5). Longitudinal sections through the flagellum and flagellar pocket of promastigotes, haptomonads and amastigotes showed a basal plate and central pair, while longitudinal sections and Markham rotational averaging showed presence of the central pair, radial spokes and inner and outer dynein arms in all three life stages (Suppl. Fig. 5). RNA-seq data confirmed this result, with central pair, radial spoke and dynein arm light and intermediate chains not significantly regulated between promastigotes and haptomonads (SI Datasets S9). Comparative genomics revealed conservation of essentially all major axonemal components in all species analyzed (with the exception of *Perkinsela* which has lost its flagellum), providing no putative markers for the capacity for 9v axoneme restructuring. RNA-seq showed no large changes in mRNA levels of axonemal proteins between promastigotes and haptomonads, consistent with this result (SI Datasets S9).

The PFR is a large para-axonemal structure required for motility. It is normally present in all trypanosomatid flagella, but is usually lost in amastigotes. We asked whether the PFR is present in different developmental stages, and whether it undergoes restructuring to form the haptomonad adhesive pad. TEM demonstrated that the PFR is present in promastigotes, but was shortened in mature haptomonads and amastigotes (Figs. 5A-C). Immunofluorescence using an antibody recognizing PFR2, a major PFR component, showed a similar flagellar localization of PFR to that of the *Trypanosoma* and *Leishmania* promastigotes, and uneven loss of the PFR in haptomonads (Fig. 5D). The characteristic quasi-crystaline structure of the PFR was only present next to the axoneme in the promastigote bulge and haptomonad attachment plaque (when visible), suggesting that elaboration of the flagellum does not correlate with expansion of the PFR. RNA-seq also showed similar PFR mRNA levels between promastigotes and haptomonads (SI Datasets S9). Comparative genomics revealed that *Bodo* and *Paratrypanosoma* possess almost all known PFR components, indicating greater conservation of the PFR than the FAZ.

Promastigote to haptomonad interconversion was modulated by the growth medium, suggesting links between the morphological adaptation and energy source. We analyzed which metabolic pathways are likely active in *Paratrypanosoma,* and which were lost early in the evolution of trypanosomatids (SI Datasets S10). When compared to *B. saltans*, *Paratrypanosoma* shows a loss of many proteases, peptidases, cathepsins and various enzymes needed for hydrolysis of complex sugars. This suggests a loss of metabolic enzymes for digestion of complex energy sources early in the evolution of parasitism. With the exception of xanthine-guanine phosphoribosyltransferase, *Paratrypanosoma* encodes all components of the purine salvage mechanism, which is thus likely operational (SI Datasets S10). However, it lacks arginases needed for the urea cycle and has lost ornithine aminotransferases and xanthine dehydrogenases, thus rendering the urea cycle and oxidative metabolism of purines non-functional (SI Datasets S10). Concerning lipid metabolism, it possesses the methylmalonyl pathway, which converts propionyl-CoA, a product of odd chain fatty acid oxidization, into succinyl-CoA. This pathway has been lost in the salivarian trypanosomes and *Phytomonas* (SI Datasets S10). *Paratrypanosoma* also has all the enzymes needed for ether-lipid biosynthesis except 1-acyl-sn-glycerol-3-phosphate acyltransferase, which is responsible for the second acyltransferase reaction of phosphatidic acid formation. It also encodes a pathway needed for phosphonolipid formation carried out by phosphoenolpyruvate mutase, previously identified only in *B. saltans* and *T. cruzi*. However, ATP citrate lyase and synthase are absent from all trypanosomatids including *Paratrypanosoma*, implying that they are unable to convert mitochondrial acetyl-CoA to citrate. Only *Paratrypanosoma* and Leishmaniinae are able to convert and subsequently oxidize methionine into succinyl-CoA. Finally, the tryptophan degradation pathway is present in *N. gruberi, T. borreli, B. saltans,* as well as in *Paratrypanosoma*, but lost from all other trypanosomatids (SI Datasets S10).

DISCUSSION

We have demonstrated that *Paratrypanosoma*, the most basal-branching trypanosomatid derived from free-living bodonids (3), assumes three different morphotypes characteristic for trypanosomatids. In a liquid cultivation medium, it alternates between motile promastigote and surface-attached haptomonad, both capable of division. Transfer of the culture onto an agar plate triggers transformation into yet another distinct morphotype, amastigote. The fact that these morphotypes are common among mono- and dixenous trypanosomatids indicates that the ancestral trypanosomatid was likely endowed with morphological

flexibility, advantageous when faced with dramatically different conditions during the evolution of parasitism of invertebrate and vertebrate hosts. Therefore, the trait of extensive inter-stage morphological transformation of *Leishmania* and *Trypanosoma* in their mammalian and insect hosts seems to have already existed in their monoxenous predecessor. Consequently, the wide array of trypanosomatid morphotypes (13) likely did not originate within the context of dixenous parasitism, but represents yet another character state that predated the two-host life style (23).

The capacity to firmly but transiently attach to a substrate by the flagellum, apparently preventing their discharge from the host, has been described in trypanosomatids, as well as in bodonids (24). This feature might have predisposed trypanosomatids for their highly successful initial radiation in insects (1), which required fixation to the host's gut. Interestingly, the attachment associated with extensive remodeling of the flagellum, as in other trypanosomatids (25), seems to be a central feature in the life cycle of *Paratrypanosoma*. Attachment and flagellum shortening could occur without a division event, while we only observed flagellum growth and detachment following division of a haptomonad. This has similarities to both *Leishmania*, in which flagellum shortening can occur without division, and *T. brucei*, in which division to generate dissimilar daughters is used for life cycle stage transition. Remodeling of the flagellum in haptomonads involved expansion of flagellum/cell attachment in the distal pocket region, suggesting the FAZ proteins may contribute to surface attachment.

Submission PDF

OG gains and losses indicate *Paratrypanosoma* has diverged significantly from common ancestor of trypanosomatids (Suppl. Fig. 6), however, it seems to retain more ancestral features than other trypanosomatid lineages except the stercorarian trypanosomes (Fig. 3). Recent comparison of the *B. saltans* and trypanosomatid genomes revealed that metabolic losses accompanied the emergence of obligatory parasitism (4) and no further gene loss or streamlining of the genome occurred in the crown trypanosomatids (5, 23). Indeed, our analyses revealed a massive loss of proteases, peptidases and cathepsins involved in the breakdown of polypeptides. *Paratrypanosoma* and other trypanosomatids have also lost the receptor-mediated endocytosis of macromolecules, cobalamin biosynthesis, and lysosomal pro-X exopeptidase and ammonium transporter, compelling trypanosomatids into foraging nitrogen from other sources. *Paratrypanosoma* has gained or expanded several gene families that were not present in its free-living predecessor, including transmembrane transporters for scavenging amino acids and other metabolites from the host (4, 5, 19). A particularly interesting example of gene families that likely underwent expansion early in the evolution of trypanosomatids is the dispersed gene family found in *Paratrypanosoma* and the stercorarian trypanosomes only. These abundant secreted proteins may play a role in host-parasite interactions (22, 25).

The ultrastructure of *Paratrypanosoma* includes a *Leishmania*-like FAZ including a microtubule quartet and a *T. cruzi*-like cytostome. This supports the hypothesis that the extended FAZ evolved once in the *Trypanosoma* lineage, and indicates that the cytostome was an ancestral feature, retained in *T. cruzi* but lost in *T. brucei*, *Leishmania*, *Phytomonas* and several monoxenous lineages analyzed thus far (Fig. 6). This implies that there has been a streamlining of ultrastructure analogous to the streamlining of the genome, with the notable exception of innovation to generate an extended FAZ among *Trypanosoma* species.

The differences in transcriptome of promastigotes and haptomonads co-existing in culture are comparable in magnitude to the differences between *L. mexicana* amastigotes (mammalian host) and promastigotes (sandfly vectors) (19, 26). Our analysis of the morphologies *Paratrypanosoma* can attain in culture, its genome and transcription profile and the ultrastructure of the flagellar pocket/cytostome complex has uncovered features likely present in the ancestors of the three human-infective trypanosomatid lineages (Fig. 6). Future studies of this very interesting protist will be particularly informative in regard to how trypanosomatid parasites have evolved from the free-living bodonids.

MATERIALS AND METHODS

Materials and Methods and associated references are available in SI Appendix, Materials and Methods.

ACKNOWLEDGEMENTS

We thank Milena Svobodová (Charles University) for providing the studied strain, Derrick Robinson (Université de Bordeaux) and Keith Gull (Oxford University) for providing antibodies, Pavel Hrouzek (Institute of Microbiology, Třeboň) for help with time-lapse video, Anzhelika Butenko (University of Ostrava) for help with bioinformatics, and Tomáš Bílý (Biology Centre) for help with microscopy. Support from the Czech Grant Agency (14-23986S to J.L., 17-10656S to V.Y. and J.V.), and ERC CZ award to J.L. is kindly acknowledged.

1. Maslov DA, Votýpka J, Yurchenko V, Lukeš J (2013) Diversity and phylogeny of insect trypanosomatids: all that is hidden shall be revealed. *Trends Parasitol* 29:43-52.
2. Yurchenko V, et al. (2014) Ultrastructure and molecular phylogeny of four new species of monoxenous trypanosomatids from flies (Diptera: Brachycera) with redefinition of the genus Wallaceina. *Folia parasit* 61(2):97-112.
3. Flegontov P, et al. (2013) *Paratrypanosoma* is a novel early-branching trypanosomatid. *Curr Biol* 23:1787-1793.
4. Opperdoes FR, Butenko A, Flegontov P, Yurchenko V, Lukeš J (2016) Comparative metabolism of free-living *Bodo saltans* and parasitic trypanosomatids. *J Eukaryot Microbiol* 63:657-678.
5. Jackson AP, et al. (2016) Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism. *Curr Biol* 26:161-172.
6. Ginger ML, Portman N, McKean PG (2008) Swimming with protists: perception, motility and flagellum assembly. *Nat Rev Microbiol* 6:838-850.
7. Wheeler RJ, Gluenz E, Gull K (2015) Basal body multipotency and axonemal remodelling are two pathways to a 9+0 flagellum. *Nat Commun* 6:8964.
8. Gluenz E, Wheeler RJ, Hughes L, Vaughan S (2015) Scanning and three-dimensional electron microscopy methods for the study of *Trypanosoma brucei* and *Leishmania mexicana* flagella. *Meth Cell Biol* 127:509-542.
9. Engstler M, et al. (2007) Hydrodynamic flow-mediated protein sorting on the cell surface of trypanosomes. *Cell* 131:505-515.
10. Broadhead R, et al. (2006) Flagellar motility is required for the viability of the bloodstream trypanosome. *Nature* 440:224-227.
11. Szempruch AJ, et al. (2016) Extracellular vesicles from *Trypanosoma brucei* mediate virulence factor transfer and cause host anemia. *Cell* 164:246-257.
12. Imhof S, et al. (2016) Flagellar membrane fusion and protein exchange in trypanosomes; a new form of cell-cell communication? *F1000Res* 5:682.
13. Wheeler RJ, Gluenz E, Gull K (2013) The limits on trypanosomatid morphological diversity. *PLoS One* 8:e79581.
14. Sunter JD, et al. (2015) Modulation of flagellum attachment zone protein FLAM3 and regulation of the cell shape in *Trypanosoma brucei* life cycle transitions. *J Cell Sci* 128:3117-3130.
15. Beattie P, Gull K (1997) Cytoskeletal architecture and components involved in the attachment of *Trypanosoma congolense* epimastigotes. *Parasitology* 115:47-55.
16. Vickerman K, Tetley L (1990) Flagellar surfaces of parasitic protozoa and their role in attachment. *Ciliary and Flagellar Membranes*, ed Bloodgood RA (Booknews, London), pp 267-304.
17. Wakid MH, Bates PA (2004) Flagellar attachment of *Leishmania* promastigotes to plastic film *in vitro*. *Exp Parasitol* 106:173-178.
18. del Pilar Corena M, et al. (2005) Carbonic anhydrase in the adult mosquito midgut. *J Exp Biol* 208:3263-3273.
19. Flegontov P, et al. (2016) Genome of *Leptomonas pyrrhocoris*: a high-quality reference for monoxenous trypanosomatids and new insights into evolution of *Leishmania*. *Sci Rep* 6:23704.
20. Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM (2015) BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* 31:3210-3212.
21. Emms DM, Kelly S (2015) OrthoFinder: solving fundamental biases in whole genome comparisons dramatically improves orthogroup inference accuracy. *Genome Biol* 16:157.
22. Lander N, et al. (2010) Localization and developmental regulation of a dispersed gene family 1 protein in *Trypanosoma cruzi*. *Infect Immun* 78:231-240.
23. Janouškovec J, Keeling PJ (2016) Evolution: causality and the origin of parasitism. *Curr Biol* 26:R174-177.
24. Gadelha C, Wickstead B, de Souza W, Gull K, Cunha-e-Silva N (2005) Cryptic paraflagellar rod in endosymbiont-containing kinetoplastid protozoa. *Eukaryot Cell* 4:516-525.
25. Lacomble S, Portman N, Gull K (2009) A protein-protein interaction map of the *Trypanosoma brucei* paraflagellar rod. *PLoS One* 4:e7685.
26. Fiebig M, Kelly S, Gluenz E (2015) Comparative life cycle transcriptomics revises *Leishmania mexicana* genome annotation and links a chromosome duplication with parasitism of vertebrates. *PLoS Pathog* 11:e1005186.

SUPPLEMENTARY MATERIALS AND METHODS

Paratrypanosoma cultivation and sample preparation for electron microscopy

The axenic culture was maintained at 27°C in mix of RPMI 1640 and M199 cultivation media (1:1) at pH 7.0 with addition of 10% (v/v) fetal calf serum, 2% sterile human urine, 10 µg/ml of hemin and penicillin-streptomycin. Amastigotes were generated by culture on semi-solid agarose plates. Autoclaved 3% [w/v] agarose was cooled to 65°C, and diluted ten times in a pre-warmed cultivation medium. The resulting 0.3% agarose solution was poured into Petri dishes (85 x 15 mm), dried for 1 hr in a laminar flow hood, and inoculated with $4x10^5$ or $2x10^6$ resuspended log phase cells. Light and electron microscopy (EM) were performed as described previously (2). For scanning EM, haptomonads were grown and fixed directly on cover slips, while for high pressure freezing transmission EM (2), they were prepared using propylene-oxide treatment to dissolve the plastic substrate as described elsewhere (15). Electron tomography was performed as described previously (8).

Time-lapse videos

Promastigote/haptomonad interconversion was analyzed by automated time-lapse videomicroscopy. Log phase ($1x10^6$ cells/ml) promastigote culture was transferred into a 35 mm µ-dish with glass bottom (iBidi) and recorded for 24 hrs at 20°C in a chamber with controlled temperature.

Experimental infection of mosquitoes and mice

Two different ways of infecting laboratory-reared mosquitoes (*Culex quinquefasciatus*) were tested. Mosquitoes were starved for 24 hrs and then fed for 2 hrs on a cotton pellet pre-soaked in a 10% sugar solution with *Paratrypanosoma*, which was shown to survive in this solution for up to 2 days. Two independent experiments each with 50 mosquitoes were performed with 10^6 and 10^7 late log cells. Forty mosquito females were infected also by feeding through a chick-skin membrane on suspension of parasites mixed 1:10 with heat-inactivated rabbit blood (final concentration was 10^7 late log cells/ml). Engorged mosquitoes were separated and maintained in the appropriate conditions (23°C, 80% humidity, 12 hrs day/light). The presence of *Paratrypanosoma* in mosquito intestine was checked on days 1, 2, 3, 4, 6, and 14 post-infection by dissection of 4 to 10 specimens at each time point. Four laboratory BALB/c mice were intraperitoneally and subcutaneously injected with 10^7 late-log stage cells. The course of infection was recorded weekly for one month (the mice were bled from the tail).

Attachment plaque digestion

To assess attachment plaque resistance to digestive enzymes, haptomonad culture was incubated with various enzymes for 3 hrs with regular intense shaking (see SI Datasets S1), and detached cells were counted. For immunofluorescence, $1x10^7$ pelleted cells were fixed for 10 min with 4% (w/v) paraformaldehyde in phosphate-buffered solution (PBS), rinsed in PBS, allowed to settle on a poly-*L*-lysine coated slide, kept for 1 hr in a blocking buffer containing 5% [w/v] nonfat dry milk in PBS, and for 2 hrs incubated with one of the following primary antibodies: anti-PFR2 (24), or anti-DOT1. After their removal, the cells were incubated for 1 hr in blocking buffer with Alexa Fluor® 488-conjugated secondary antibody, rinsed with PBS, mounted in DAPI-containing ProLong diamond antifade reagent and observed.

Genome and transcriptome assembly and annotation

Genome assembly was made with GS De Novo Assembler (Newbler) v2.9 using reads obtained on the Illumina MiSeq platform combining 3,4 million reads from a paired-end library (insert size 0.4 kb; average read length upon trimming 241 nt; 29x coverage) and 4 million reads from a mate-pair library (insert size 1-8 kb; average read length upon trimming 255 nt; 34x coverage). Total RNA and poly(A)-enriched fractions were used to prepare MiSeq paired-end libraries (insert size 0.28 kb; read length 150 nt) to sequence the transcriptome with the Illumina MiSeq system, resulting in 34 and 37 million high quality reads, respectively. All sequencing data were deposited to the TriTrypDB (http://tritrypdb.org/tritrypdb/). Transcriptomic reads were aligned to the genome assembly using Bowtie2 v. 2.2.5 with the '--end-to-end' and '--very-sensitive' options. Augustus v2.5.5 was used for annotation and its accuracy was improved by retraining with a set of 100 highly conserved gene models. The annotation was further manually enhanced as follows: transcribed ORFs longer than 100 amino acids, not predicted by Augustus, were added to the annotation and gene models with start sites predicted in regions with no transcription were corrected based on RNA-seq data. Subsequently, Blast2GO programs were used to obtain functional gene annotations. To analyze differential gene expression, three independent replicates of transcriptomic Illumina HiSeq libraries from both sessilemastigotes and promastigotes were generated. Differential gene expression analysis of six libraries with ~50 million reads each (insert size 280 bp; read length 101nt) was performed using a procedure described elsewhere (19).

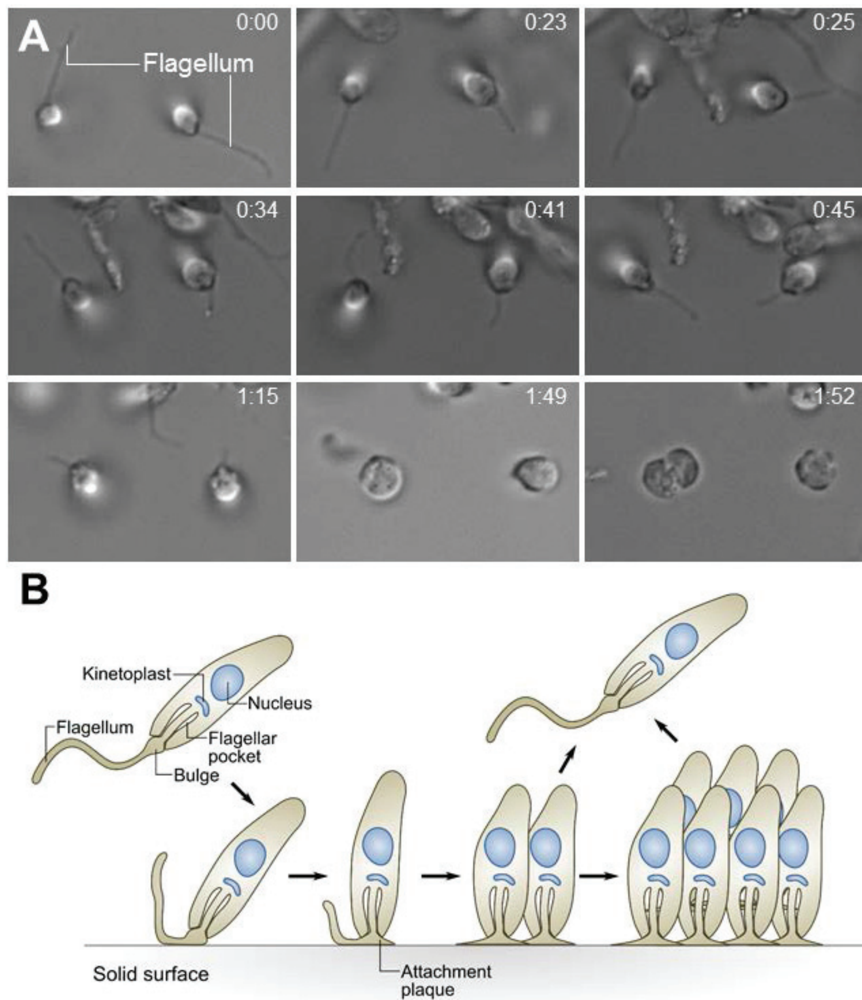
Gene family, gene ontology and differential expression analyses

Orthologous groups (OGs) were inferred using Orthofinder software v.0.61 (21). Annotated proteins of 29 kinetoplastid species and a heterolobosean outgroup (SI Datasets S2) downloaded from TriTrypDB v. 31, Marine Microbial Eukaryote Transcriptome Sequencing Project (http://marinemicroeukaryotes.org/), and Welcome Trust Sanger Institute (http://www.sanger.ac.uk/resources/downloads/protozoa/) were combined with newly annotated proteins of *Paratrypanosoma* and *Blechomonas ayalai.* The Count program was subsequently used for mapping gene family gains/losses (the Dollo parsimony algorithm) onto a reference species cladogram based on a multigene phylogenetic tree. Each protein, from set of 98 unique proteins present in all species, was separately aligned using mafft and informative positions with consecutive concatenation was performed using Gblocks. Multiprotein tree was constructed using RAxML v. 8.2.1 (LG+Γ model) with 1,000 bootstraps and Phylobayes v. 4.1c (GTR+ Γ+CAT model) running 8 independent chains for 10,000 cycles.

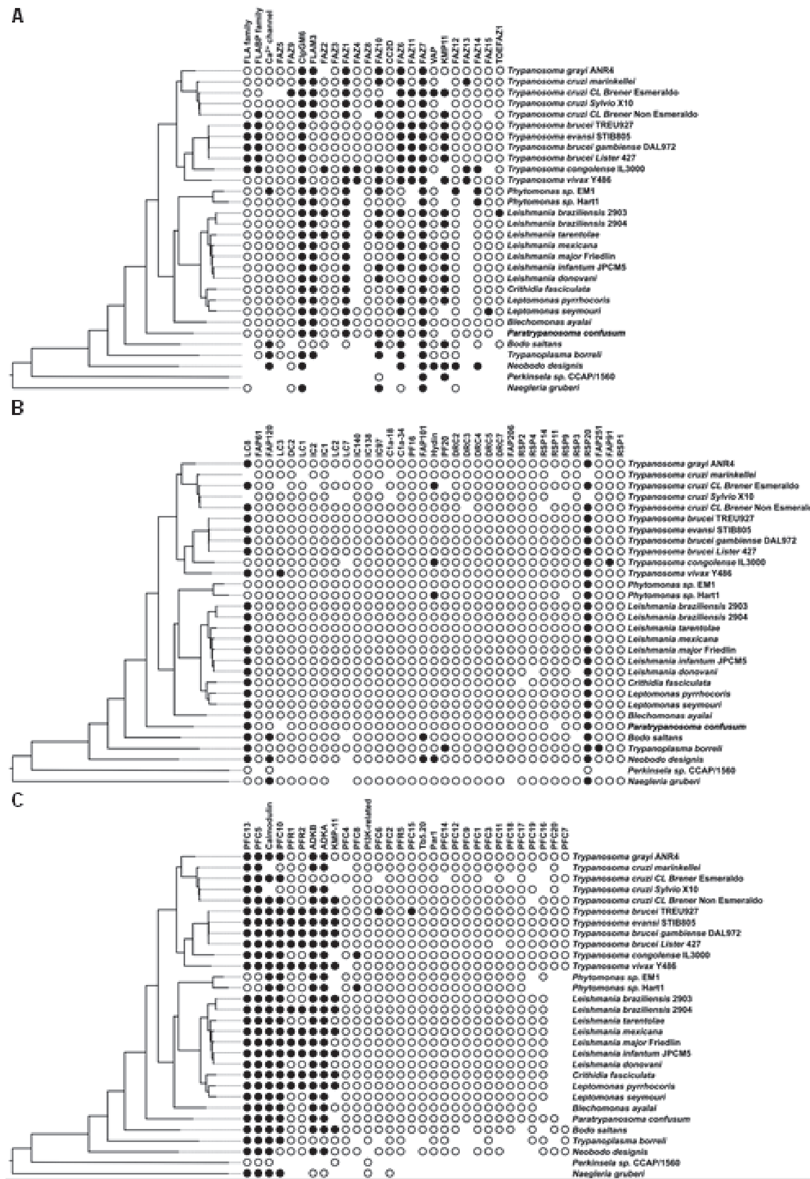
Gene ontology (GO) annotation of *Paratrypanosoma* gene families gained or lost at selected nodes was performed using the Blast2GO software with the following settings: BLASTP (10^{-10} E-value cut-off) was run using Blast2GO's CloudBlast service, retaining best 20 hits and filtering low complexity regions. Mapping GO terms onto Blast hits was followed by selection of most specific GO terms (an annotation cut-off of 55 was used). Resulting annotations were visualized by generation of GO graphs and multi-level pie charts for each GO term category (cellular component, biological process and molecular function). A differential expression analysis using transcriptomic data from both promastigotes and haptomonads was conducted in CLC Genomic Workbench v8 and only genes with an expression fold change ≥ 2 and an FDR-corrected p-value ≤ 0.05 were analyzed further. GO terms enrichment was also analyzed for protein-coding genes significantly over/under

expressed in haptomonads as compared to promastigotes, vs. all protein-coding genes using Fisher's exact test with a FDR corrected p-value cut-off of 0.05.

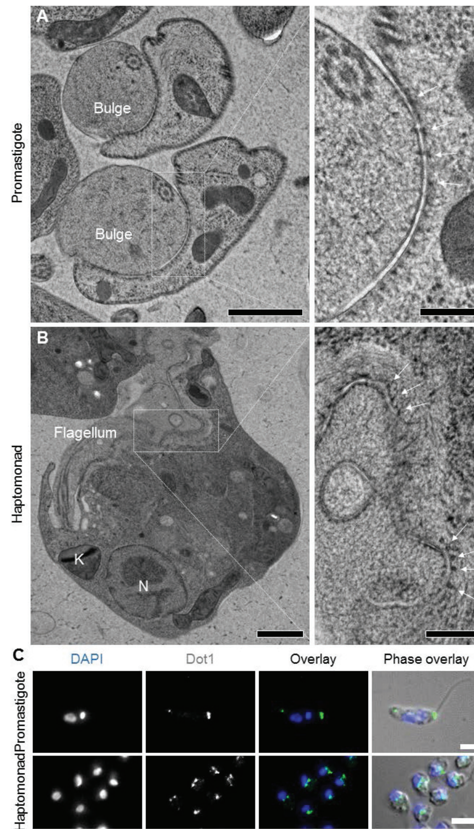
Supplementary Figures



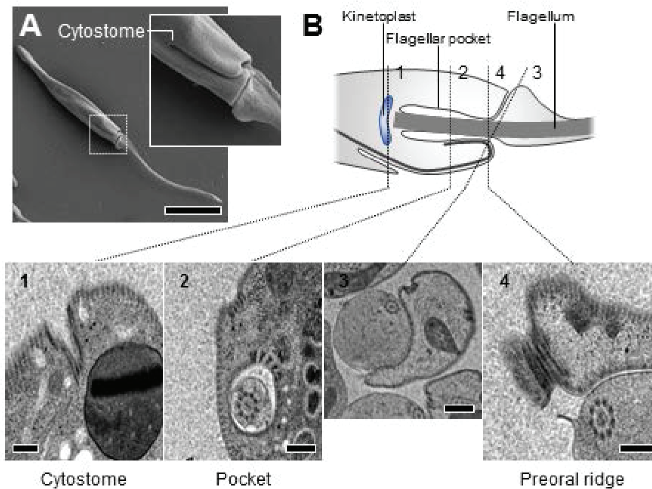
Suppl. Fig. 1. Cartoon of the attachment process. The promastigote freely swims and can attach to a surface by the bulge. The flagellum shortens and the cell assumes an upright position with the bulge expanding into a thin attachment pad. Division events can generate cells which either attach to the surface or grow a long flagellum and leave the colony.



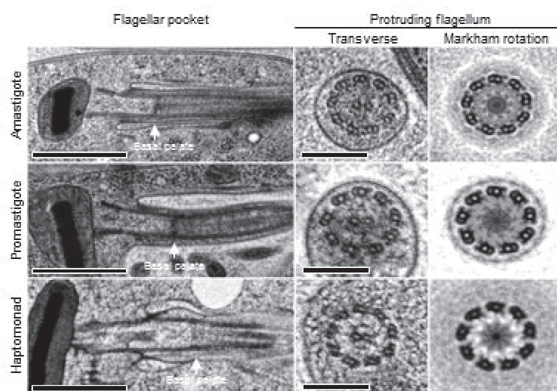
Suppl. Fig. 2. Conservation of cytoskeletal proteins across the trypanosomatids. Matrix summarizing presence of a single (open circle), multiple (filled circle) or no (space) orthologs of a gene family for **A)** FAZ, **B)** axoneme and **C)** PFR genes.



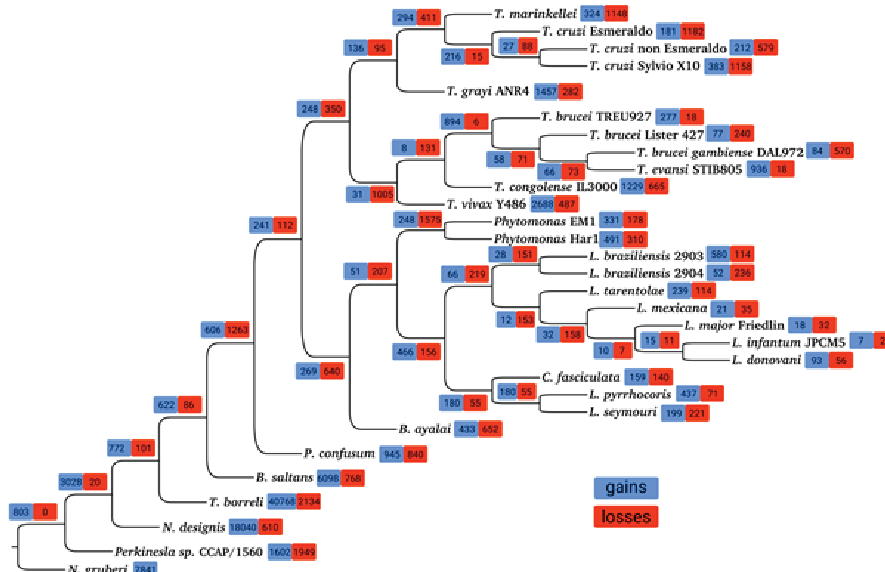
Suppl. Fig. 3. The flagellar bulge is elaborated in haptomonads for surface attachment. **A)** Detail of attachment between the flagellum/bulge with the cell body in the promastigote. Desmosome-like attachments, similar to the *T. brucei* and *Leishmania* FAZ, are indicated with arrows. **B)** Detail of attachment between the flagellum/bulge with the haptomonad cell body. **C)** Immunofluorescence using Dot1, a *T. brucei* anti-FAZ antibody, of promastigotes and haptomonads. Dot1 labels a distal ring or horseshoe structure in promastigotes, similar to the localization of FAZ10 in *Leishmania*. The localization is elaborated in haptomonads. Scale bars indicate A) 1 µm, detail 250nm B) 1 µm, detail 250nm C) 2µm



Suppl. Fig. 4 *Paratrypanosoma* has a cytostome similar to *T. cruzi* and a pocket architecture similar to *Leishmania*. A) SEM of the promastigote showing an indentation, similar to an extended *T. cruzi* preoral ridge, extending from the flagellar pocket to the cytostome. **B)** Cartoon of the pocket/cytostome of *Paratrypanosoma*, based on TEM of the flagellar pocket. A set of highly decorated microtubules runs from inside the pocket (15), around the end of the flagellar pocket neck (8) to the cell pellicle and back to the cytostome. Scale bars indicate A) 5µm, detail is 3.3 times enlarged B1,2,4) 250nm B3) 500nm



Suppl. Fig. 5. Restructuring of the *Paratrypanosoma* flagellum between promastigotes, haptomonads and amastigotes. TEM of longitudinal sections through the flagellar pocket and transverse sections through the protruding portion of the flagellum in the promastigote, haptomonad and amastigote. Ninefold Markham rotational averaging of axoneme structure shows presence of inner and outer dynein arms and radial spokes in the amastigote, promastigote and haptomonad. White arrows indicate the basal plate. Scale bars indicate 500nm and flagellar details 200nm.



Suppl. Fig. 6. Gene family gains/losses mapped on cladogram using Dollo parsimony. Gained and lost gene families (GOs) are mapped on cladogram which topology was obtained from multiprotein phylogenetic analysis. Gene families gained are labeled in blue and gene families that were lost in red, respectively.

All SI datasets and Suppl. Videos can be downloaded here: https://github.com/tskalicky/paratrypanosoma

SI Datasets S1. (**A**) **List of enzymes used to digest the attachment pad of haptomonads kept for 24 hrs in the cultivation medium**. Enzymes are listed according to their activity. For some enzymes (marked with an asterisk) the concentration is listed in units. (**B**) List of enzymes used to digest the attachment pad of haptomonads kept for 3 hrs in 1x PBS. Enzymes are listed according to their activity.

SI Datasets S2. Phyletic patterns, differential expression data and annotations for all *Paratrypanosoma* genes. Expression measured in RPKM at both life cycle stages, fold change and FDR-corrected *p*-value are shown for each gene.

SI Datasets S3. Gene ontology terms significantly enriched in gene sets upregulated in haptomonads and promastigotes. A) Gene ontology (GO) terms significantly enriched (FDR-corrected *p*-value < 0.05) in haptomonads. **B)** GO terms significantly enriched in promastigotes.

SI Datasets S4. Genomes used for gene family analyses. Genomes were obtained from publicly available databases such as Welcome Trust, GenBank, TriTrypDB (version 31) or Marine Microbial Eukaryote Transcriptome Sequencing Project (MMETSP).

SI Datasets S5. Annotations of ancestral genes retained in *Paratrypanosoma* and lost in the other trypanosomatids. *Paratrypanosoma* genes belonging to orthologous groups with this phyletic pattern are shown. Annotations were generated using Blast2GO. Metabolic proteins are highlighted.

SI Datasets S6. Annotations of ancestral genes retained in *Paratrypanosoma* and stercorarian trypanosomes and lost in the other trypanosomatids. *Paratrypanosoma* genes belonging to orthologous groups with this phyletic pattern are shown. Annotation were generated using Blast2GO. Metabolic proteins are highlighted, as well as proteins of the dispersed gene family.

SI Datasets S7. Relationship of phyletic patterns and stage-specific gene expression in *Paratrypanosoma*. One-way ANOVA analysis combined with Tukey's honest significance test shows that differential gene expression at the haptomonad and promastigote stages depends on phyletic patterns of corresponding orthologous groups (p-value = 1.22×10^{-7}). The presence/absence patterns across six species groups (phyletic patterns) are shown in the left-most rows and on top of the matrix. Genes were grouped by phyletic patterns, and for each group average fold change was calculated. Positive fold change corresponds to upregulation in haptomonads, and negative fold change to upregulation in promastigotes. Number of genes with significant stage-specific differences in expression is also shown for each group. ANOVA p-values corrected for multiple testing appear in each cell of the matrix and represent pairwise comparisons of gene groups.

SI Datasets S8. Genes belonging to orthologous groups gained at the *Trypanosoma* node.

Genes gained at *Trypanosoma* node were identified by Count program, applying Dollo parsimony analysis on the orthologous groups created by Orthofinder. Appropriate proteins were extracted from working dataset of proteoms and blasted in Blast2GO against nr protein database with default settings. Column labels describe to which orthologous group proteins belong, their gene ID, description, protein length, number of blast hits, E-value and avg. % similarity.

SI Datasets S9. Differential expression of cytoskeletal proteins in haptomonads and promastigotes. Genes significantly upregulated in haptomonads (SI Datasets S2) are marked in green color. Multi-gene families are also highlighted with various colors.

SI Datasets S10. Presence/absence of selected metabolic enzymes in all analyzed species. Presence is marked with green color and absence in red, respectively. Numbers represent number of gene copies in each species. Abbreviations explanation: Baya = *Blechomonas ayalai,* Bsal = *Bodo saltans*, Cfas = *Crithidia fasciculata*, Lbr1 = *Leishmania braziliensis* MHOM/BR/75/M2903, Lbr2 = *Leishmania braziliensis* MHOM/BR/75/M2904, Ldon = *Leishmania donovani*, Linf = *Leishmania infantum,* Lmaj = *Leishmania major* Friedlin*,* Lmex = *Leishmania mexicana*, Ltar = *Leishmania tarentolae*, Lpyr = *Leptomonas pyrrhocoris*, Lsey = *Leptomonas seymouri,* Ndesig = *Neobodo designis*, Ngrub = *Naegleria gruberi, Pconf = Paratrypanosoma confusum*, Pem1 = *Phytomonas* sp. EM1*,* Perk = *Perkinsela* sp. CCAP-1560, Phar = *Phytomonas* sp. HART1, Tbor = *Trypanoplasma borreli,* Tbr1 = *Trypanosoma brucei*

Lister 427, Tbrg = *Trypanosoma brucei gambiense* DAL972, Tbru = *Trypanosoma brucei* TREU927, Tcon = *Trypanosoma congolense,* Tcr1 = *Trypanosoma cruzi* CL Brener Esmeraldo-like, Tcr2 = *Trypanosoma cruzi* CL Brener Non-Esmeraldo-like, Tcr5 = *Trypanosoma cruzi* Sylvio, Teva = *Trypanosoma evansi,* Tgra = *Trypanosoma grayi*, Tmar = *Trypanosoma cruzi marinkellei*, Tviv = *Trypanosoma vivax*.

Suppl. Video 1. Time-lapse video showing promastigote cell transformation into haptomonad, multiple divisions and recreation of motile promastigote cells after division. Video shows time span of about 5 and half hour of real time.

Suppl. Video 2. Time-lapse video showing promastigote cell transformation into haptomonad with focus on cell flagellum and cell division. Video shows time span of about 5 hours of real time.

6 Curriculum vitae

RNDr. Tomáš Skalický

Date of birth: 28 August 1985

Nationality: Czech

Home address: Závořická 555, 789 69 Postřelmov, Czech Republic
Telephone number: +420 739 22 11 62
E.mail: skalicky@paru.cas.cz, tomas.skalicky@seznam.cz

EDUCATION

2011 – present **Ph.D. student** of Molecular and Cell Biology and Genetics

Department of Molecular Biology and Genetics, Faculty of Science, **University of South Bohemia**, České Budějovice

Institute of Parasitology, **Czech Academy of Sciences,** Czech Republic

Thesis: Insight to insect trypanosomatid diversity via whole genome sequencing

Supervisor: Julius Lukeš

2013 **RNDr., Molecular Biology and Genetics**

Faculty of Science, **University of South Bohemia**, Czech Republic

2009 – 2011 **M.S., Experimental Biology**

Department of Molecular Biology, Faculty of Science, **University of South Bohemia**, Czech Republic. Thesis: Functional analysis of Ssc and Iba57 proteins in *Trypanosoma brucei*. Supervisor: Julius Lukeš

2005 – 2009 **B.S., Biology**

Department of Genetics, Faculty of Science, **University of South Bohemia**, Czech Republic. Thesis: Effect of selected exogenous and endogenous factors on lineseed (*Linum usitatissimum L.*) transformation efficiency. Supervisor: Slavomír Rakouský

PROFESSIONAL EXPERIENCE

EMPLOYMENT:

2011 – present	**Graduate student,** Institute of Parasitology, **Biology Centre of the Czech Academy of Sciences**
2010 – 2011	**Research worker**, Institute of Parasitology, **Biology Centre of the Czech Academy of Sciences**
2007 – 2009	**Research worker**, Faculty of Health & Social Studies, **University of South Bohemia**, České Budějovice

RESEARCH STAYS:

2016	Visiting Student (XI.-XII.) Matthew Berriman, Parasite Genomics group, **Welcome Trust Sanger Institute**, Hinxton, United Kingdom

---

PEER-REVIEWED PUBLICATIONS

Číčová Z, Dejung M, **Skalický T**, Eisenhuth N, Hanselmann S, Morriswood B, Figueiredo LM, Butter F, Janzen CJ (2016) Two flagellar BAR domain proteins in *Trypanosoma brucei* with stage-specific regulation. **Scientific Reports** 6:35826 (IF = 5.228).

Drini S, Criscuolo A, Lechat P, Imamura H, **Skalický T**, Rachidi N, Lukeš J, Dujardin JC, Späth GF (2016) Species- and Strain-Specific Adaptation of the HSP70 Super Family in Pathogenic Trypanosomatids. **Genome Biology and Evolution** 8(6):1980-95 (IF = 4.098).

Dobáková E, Flegontov P, **Skalický T**, Lukeš J (2015) Unexpectedly Streamlined Mitochondrial Genome of the Euglenozoan *Euglena gracilis*. **Genome Biology and Evolution** 7(12):3358-67 (IF = 4.098).

Votýpka J, Rádrová J, **Skalický T**, Jirků M, Jirsová D, Mihalca AD, D'Amico G, Petrželková KJ, Modrý D, Lukeš J (2015) A tsetse and tabanid fly survey of African great apes habitats reveals the presence of a novel trypanosome lineage but the absence of *Trypanosoma brucei*. **International Journal of Parasitology** 45(12):741-8 (IF = 4.242).

Škodová I, Verner Z, **Skalický T**, Votýpka J, Horváth A, Lukěš J (2015) Lineage-specific activities of a multipotent mitochondrion of trypanosomatid flagellates. **Molecular and Biochemical Parasitology** 96(1):55-67 (IF = 3.761).

Lukeš J, **Skalický T**, Týč J, Votýpka J, Yurchenko V (2014) Evolution of parasitism in kinetoplastid flagellates. **Mol Biochem Parasit** 195(2):115-122 (IF = 2.068)

Flegontov P, Votýpka J, **Skalický T**, Logacheva MD, Penin AA, Tanifuji G, Onodera NT, Kondrashov AS, Volf P, Archibald JM, Lukeš J (2013) Paratrypanosoma is a novel early-branching trypanosomatid. **Curr Biol** 23(18):1787-93 (IF = 9.733).

Long S, Changmao P, Tsaousis AD, **Skalický T**, Verner Z, Wen Y, Roger AJ, Lukeš J (2011) Stage-specific Requirement for Isa1 Isa2 and proteins in the mitochondrion of *Trypanosoma brucei* and rescue the heterologous human orthologues and *Blastocystis*. **Molecular Microbiology** 81(6):1403-18 (IF = 4.347).

Beranová M, Rakouský S, Vávrová Z and **Skalický T** (2008) Sonication assisted Agrobacterium-mediated transformation enhances the transformation efficiency in flax (*Linum usitatissimum* L). **Plant Cell Tissue Organ Culture** 94:253-259 (IF = 2.286).

PRESENTATIONAS AT CONFERENCES

2017	47th International Meeting of Czech Society for Protozoology, Nové Hrady, Czech Republic. Talk presentation.
2016	46th International Meeting of Czech Society for Protozoology, Bítov, Czech Republic. Poster presentation.
	TryTax2 meeting, Ostrava, Czech Republic. Poster presentation
	Canadian Institute for Advanced Research meeting, Toronto, Canada. Talk presentation.
2015	45th International Meeting of Czech Society for Protozoology, Dubovice, Czech Republic. Poster presentation.
2014	44th International Meeting of Czech Society for Protozoology, Visalaje, Czech Republic. Poster presentation.
	ENBIK 2014, Kouty, Czech Republic. Passive attendance.

	Canadian Institute for Advanced Research meeting, Liblice u Prahy, Czech Republic. Poster presentation.
2013	43rd International Meeting of Czech Society for Protozoology, Nový Dvůr nad Vltavou, Czech Republic. Poster presentation.
	ICOP XIV International Congress of Protistology, Vancouver, Canada. Poster presentation.
2012	42nd International Meeting of Czech Society for Protozoology, Kletečná u Humpolce, Czech Republic. Talk presentation.
2010	40th International Meeting of Czech Society for Protozoology, Kouty, Czech Republic. Talk presentation.

OTHER LECTURES

2014	Illumina Days, České Budějovice, Czech Republic. Talk presentation.
2014	„Parasitological weekend for high-school teachers“ a popularization workshop. Talk presentation and practicals.

HONORS, AWARDS AND FUNDING

2013 – 2014	Grant Agency of the University of South Bohemia (reg. no. 108/2013/P, PI: Tomáš Skalický). *Insight to insect trypanosomatid diversity via whole genome sequencing.*
2013	Holtz-Conner Award from International Society of Protistologists

2010 – 2011 Premium Scholarship for outstanding study results, Faculty of Science University of South Bohemia, Czech Republic

ATTENDED WORKSHOPS

2017 OstravaPy – python workshop, Ostrava, Czech Republic

2014 Workshop on Genomics, Český Krumlov, Czech Republic

2013 Evaluation of pyrosequencing data - Bioinformatics workshop, České Budějovice, Czech Republic

2011 Workshop on Cutting-edge Bioinformatics, Shenzhen, China

SKILLS

Field work: dissection and examination of collected insects for gut parasites and their introduction into cell culture

Laboratory methods: ten years of experience with basic and advanced molecular biology methods (including genomic and transcriptomic library preparation for 454 and Illumina sequencing methods), insect trypanosomatid and *Trypnosoma brucei* cell culture cultivation, flow cytometry, various microscopy methods (ranging from light, fluorescence and confocal to transmission and scanning electron microscopy).

Bioinformatics: more than eight-year experience of linux systems, setting up and administration of linux computational servers, good knowledge of phylogenomics, genomics and transcriptomics programs, basic experience in Bash, Python and R programming.

Languages: English (full professional proficiency, TOEIC C1 level and TEOFL ITP certificates), German (limited working proficiency), Spanish (beginner), Czech (native).

PEDAGOGICAL ACTIVITY

2011 – 2012	Biology Centre of the Czech Academy of Sciences, teacher at project EKOTECH in advanced methods of molecular biology (reg.num. CZ.1.07/2.3.00/09.0200)
2011	Faculty of Science, University of South Bohemia, České Budějovice, Teaching assistant at biochemical practices

---