CellPress

Trends in Parasitology

Forum

Phylogenetic framework to explore trait evolution in Trypanosomatidae

Alexei Yu. Kostygov [1,5,*], Amanda T.S. Albanaz [1,5], Anzhelika Butenko [1,2,3], Evgeny S. Gerasimov [4], Julius Lukeš [2,3] and Vyacheslav Yurchenko [1,*]

The number of sequenced trypanosomatid genomes has reached a critical point so that they are now available for almost all genera and subgenera. Based on this, we inferred a phylogenomic tree and propose it as a framework to study trait evolution together with some examples of how to do it.

What are trypanosomatids?

The family Trypanosomatidae is one of the best studied groups of protists, encompassing flagellates that parasitize a wide range of hosts, including vertebrates, arthropods, leeches, plants, and ciliates [1]. It is mostly known, since some members cause human diseases, such as sleeping sickness (*Trypanosoma brucei*), Chagas' disease (*Trypanosoma cruzi*), and various forms of leishmaniasis (*Leishmania* spp.). According to their life cycles, these flagellates are subdivided into two nontaxonomic groups: monoxenous (with a single host) and dixenous (switching between two hosts, of which one is considered a vector). Insects are predominant hosts for the monoxenous trypanosomatids and the most frequent vectors for the dixenous ones [1].

Apart from the practical importance of some representatives, trypanosomatids share many unusual features which attract considerable research attention. These include polycistronic gene transcription and *trans*-splicing of nuclear genes, a large mass of mitochondrial DNA organized into a concatenated network, RNA editing in mitochondria, and compartmentalization of glycolysis in specialized organelles – glycosomes [2]. Some trypanosomatids harbor further remarkable features: an idiosyncratic genetic code with all three stop codons encoding amino acids, intracellular bacterial symbionts that complement missing metabolic pathways of their protistan hosts, or two joint flagella as an adaptation to life in a dynamic environment [1].

Genomics – a powerful tool to study trypanosomatid biology

Progress in the knowledge of trypanosomatid biology in the 21st century has been tightly associated with the increasing role of genomic studies. For this protistan group, the genomic era has started with the analysis of the three medically important species – *T. brucei*, *T. cruzi*, and *Leishmania major* [3]. These pioneering studies shed light on the peculiarities of trypanosomatid metabolism, cell and molecular biology, as well as numerous novel features concerning the organization of the genome. Subsequently, the genomes of many trypanosomes and leishmaniae have been analyzed with a conspicuous preponderance of species from the latter, which has a higher proportion of human pathogens [4] This allowed characterization of the dynamics of the sexual process, confirming the occurrence of intra- and interspecific hybridization, and revealing the genomic basis of the astonishing adaptability of these flagellates, which is stipulated by gene copy number variation due to the deletion or duplication of complete chromosomes or their individual segments [5].

However, for better comprehension of trypanosomes and leishmaniae, their genomic data should be compared with those of other trypanosomatids. This required expansion of genomic studies into the area of monoxenous genera, dixenous plant-parasitic *Phytomonas* spp. and the closest free-living relative of trypanosomatids – *Bodo saltans* [4,6]. Such comparative analyses highlighted extensive lifestyle-associated changes in the gene repertoire. In particular, the origin and evolution of dixeny correlated with contractions and expansions of gene families responsible for various transporters, surface molecules (mucins, *trans*-sialidases, amastins, lipophosphoglycan) and secreted proteins (peptidase gp63 and cruzipain) [4]. Meanwhile, the acquisition of ferrous iron and heme transporters apparently was a prerequisite for the origin of parasitism in the common ancestor of trypanosomatids [2]. Genome analysis for trypanosomatids bearing intracellular bacteria allowed deinterlacing complex metabolic relationships in such associations, which are characterized by a much higher biosynthetic capacity than regular (endosymbiont-free) trypanosomatids [7]. The hierarchical analysis at different taxonomic levels using comparative genomic

Box 1. Phylogenomics

Genomic sequences are valuable not only as a source of information elucidating biological properties of the studied organisms, but also as data from which their evolutionary history can be reconstructed. While phylogeny can be inferred based on individual genes, this approach has significant drawbacks. A single phylogenetic marker usually does not have enough information to resolve both close and distant relationships. This is further complicated by the discordance of evolutionary trajectories for organisms and their genes, since the latter can experience duplications and losses, introgression and horizontal transfer. The phylogenomic approach – that is, inference of phylogeny based on a substantial part of the genome (typically, multiple protein-coding genes, which preserved the single-copy state during the evolution of the group under study) – alleviates these issues. Moreover, the considerably larger amount of data allows the use of very complex substitution models with parameters estimated directly from a sequence alignment. This considerably decreases the occurrence of reconstruction artifacts as compared to the application of precomputed substitution matrices. Therefore, phylogenomic inference is preferable when a reliable and well-resolved evolutionary tree of taxa is needed.

data enabled detailed inference of gains determining organellar complexity observed in *T. brucei* [8].

Recently, the genomic sequences of at least one member of each trypanosomatid genus have become available, allowing extensive comparative analyses, which may eventually provide a comprehensive picture of this diverse group of parasites.

Trypanosomatid phylogeny: challenges and applications

The evolutionary framework is of great importance for studying different aspects of trypanosomatid biology. One of the common approaches is the analysis of trait evolution by exploring their distribution on a phylogenetic tree of studied taxa. In this respect, genomic data provide a remarkable combination, being suitable for robust phylogenetic (or more specifically, phylogenomic, see Box 1) reconstructions and providing characters that can be mapped on the inferred trees.

The analysis of trait evolution requires a phylogenetic tree with a reliable topology. However, in the absence of genomic data for many important trypanosomatid lineages, phylogenetic reconstructions revealed multiple issues, which could not be resolved using analyses based on a single gene or on a few genes. This concerns, for example, the phylogenetic position of the genera *Wallacemonas*, *Sergeia*, *Vickermania*, and *Jaenimonas*, the relationships between numerous subgenera of *Trypanosoma* and between the monoxenous genera within the subfamily Leishmaniinae. Taken together, all these uncertainties resulted in a tree with multiple polytomies [1], which preclude a proper exploration of trait evolution. Luckily, a nearly comprehensive dataset of high-quality genome sequences is now available. Below we demonstrate how it can be effectively used to address various evolutionary questions.

Updated phylogeny of trypanosomatids

Here, using a phylogenomic dataset based on publicly available data, we inferred relationships between all currently recognized trypanosomatid genera (except *Rhynchoidomonas*, which has not been re-isolated so far [1]), detailed

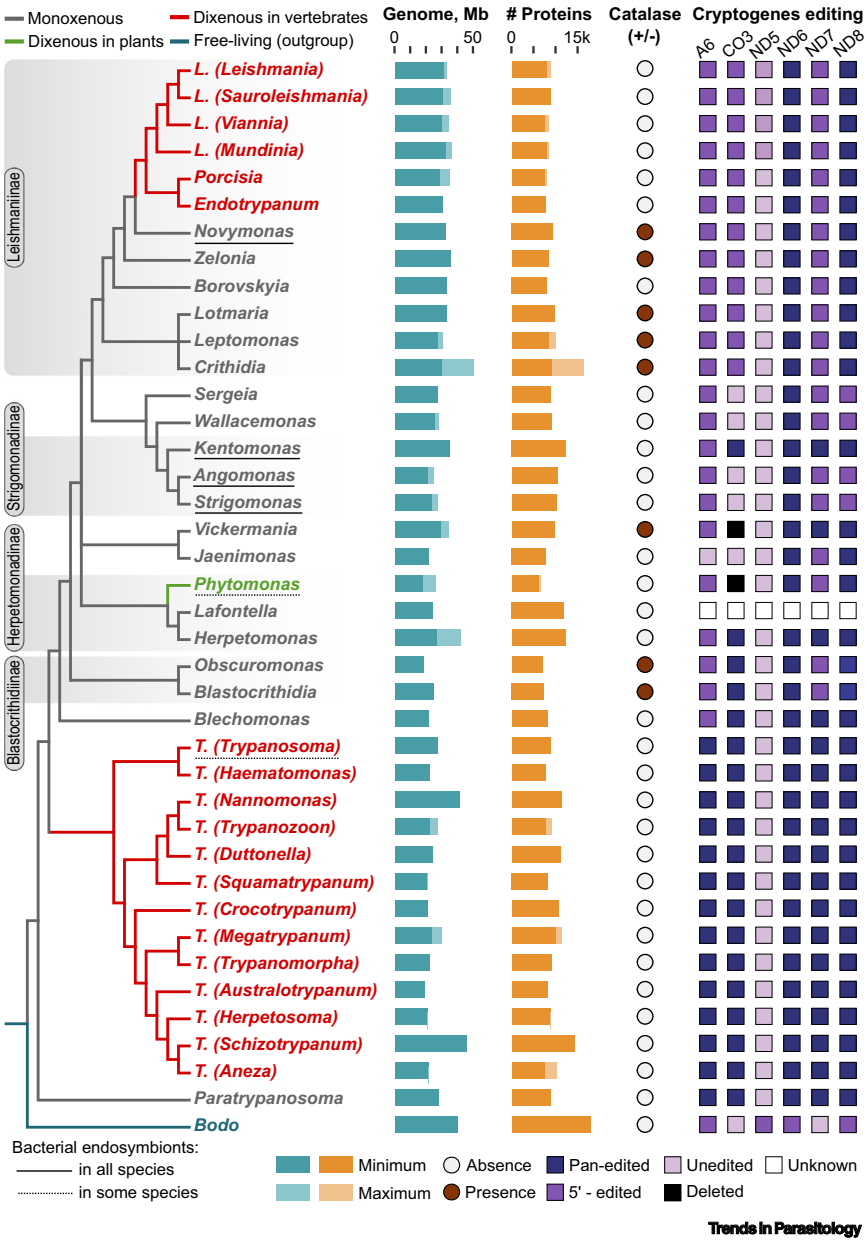

Figure 1. Phylogenomic tree of Trypanosomatidae with mapped genomic features. The tree was combined from independent analyses: at the genus level for the whole family, as well as at the species level for monoxenous Leishmaniinae, the genus *Trypanosoma* (see supplementary Figures S2 and S3) and a previously published one for the clade *Leishmania/Porcisia/Endotrypanum* [15]. Lineages are colored according to their lifestyles. Subfamilies are indicated only if they contain more than one genus. Polytomy for the genera *Crithidia*, *Leptomonas*, and *Lotmaria* reflects the inability to reliably separate them into monophyletic taxa, despite fully resolved relationships at the species level (see Box S1, Figure S1, and Tables S1–S6 in the supplemental information online).

up to subgeneric level for *Leishmania* and *Trypanosoma* (Figure 1). Although, in general, the tree agrees with previous reconstructions based on a handful of genes, there are also several important updates. One of them concerns the clade of three monoxenous genera of the subfamily Leishmaniinae – *Crithidia*, *Leptomonas*, and *Lotmaria*. A species-level phylogenomic inference (see Figures S2 and S3 in the supplemental information online) demonstrates that they cannot be separated into monophyletic taxa, thereby indicating the need for a taxonomic revision. The previously suggested relatedness of *Wallacemonas* and *Sergeia* to Strigomonadinae [9] is confirmed, but instead of being sister to each other, these two genera consecutively branch off within the common clade, which is sister to Leishmaniinae. Another important finding is the sister relationship between *Vickermania* and *Jaenimonas*, which could not be reliably inferred using 18S rRNA and gGAPDH gene sequences [10]. Our analysis also revealed the relatively early branching of the subfamily Blastocrithidiinae, which diverged fourth after *Paratrypanosoma*, *Trypanosoma*, and *Blechomonas*.

Our phylogenomic tree fully resolves the relationships between subgenera of *Trypanosoma* and confirms the previously established basal split between the aquatic and terrestrial species [11]. Interestingly, the latter group includes an early-diverging clade, in which African salivarian trypanosomes are sister to the poorly studied subgenus *Squamatrypanum*, encompassing parasites of squamates and small mammals [1]. Interestingly, the trypanosomes of Archosauria (i.e., crocodiles and birds) are not directly related, as exemplified by the relationships between the subgenera *Crocotrypanum* and *Trypanomorpha*.

Inference of trait evolution using the phylogenomic tree

Here, we illustrate with a few examples how the updated trypanosomatid tree can be used for tracing the evolution of various features in these flagellates. The integral parameters of the genomes – such as their size and coding capacity (i.e., number of encoded proteins) – demonstrate evolutionary variability with considerable differences observed, even within a genus (Figure 1).

It has been previously demonstrated that catalase, an enzyme protecting cells from the toxic hydrogen peroxide, had been acquired by three lineages – *Vickermania*, Blastocrithidiinae, and Leishmaniinae, while the dixenous members of the latter group secondarily lost it [12]. Our new inference did not reveal any additional cases of catalase acquisition, but points to its independent loss in one more member of Leishmaniinae, *Borovskyia barvae*. This might be related to the dependence of this trypanosomatid on an accompanying yeast-like fungus, which apparently possesses this enzyme [13].

Another example is the distribution of RNA editing (Box 2) for transcripts of cryptogenes in the kinetoplast genomes. It is generally accepted that the common trypanosomatid ancestor featured pan-editing for all, or the majority, of cryptogenes followed by the progressive reduction of the process in some descendants [2]. The phylogenomic tree presented here confirms this view, since the early-diverging *Paratrypanosoma* and *Trypanosoma* have pan-editing in all selected cryptogenes except *ND5*, while the switch to 5'-editing of *A6* occurred just once in the common ancestor of all other trypanosomatids. As for other cryptogenes, the reduction of the edited domain likely occurred in parallel in different lineages, which is evident from the comparison of *Kentomonas* and/or *Vickermania* with their respective relatives (Figure 1). It is also notable that the loss of *COIII* (and other subunits of cytochrome *c* oxidase complex) and *apoB* occurred independently in *Vickermania* and *Phytomonas* [14].

The well-resolved phylogeny, which for the first time encompasses almost all trypanosomatid genera and subgenera, can now be used to address a wide range of questions related to trait evolution. We believe that such studies will be further stimulated by the available large-scale genomic data accelerating the progress in our understanding of these amazing protists.

Box 2. RNA editing in trypanosomatids

The mitochondrial (kinetoplast) genome in Trypanosomatidae consists of two types of circular molecules concatenated into a single network: maxicircles and minicircles. The former carry typical mitochondrial genes coding for rRNAs and subunits of the electron transport chain. To become translatable, the primary transcripts of the protein-coding genes usually require editing, which is represented by insertions and deletions of uridines, and is executed by multiple macromolecular protein complexes and minicircle-encoded guide RNAs. Due to a significant difference between the DNA-encoded sequence and that of the mature mRNA, such genes are usually referred to as cryptogenes. The extent of editing varies among genes and trypanosomatid taxa, ranging from the pan-editing (encompassing most of the sequence), through 5'-editing (restricted to the 5' portion), to the complete loss of editing, so that a cryptogene becomes a regular gene.

Acknowledgments

The authors thank Dmitri Afonin (Moscow State University) for collecting data on RNA editing patterns. This work was primarily supported by the Czech Grant Agency grant 23-07695S to A.B. and A.Y.K. The RNA editing analysis was funded by the Russian Science Foundation grant 19-74-10008 to E.S.G. and Czech Grant Agency grant 22-01026S to J.L. and V.Y. The computational infrastructure used in analyses was purchased in the frame of the EU's Operational Program 'Just Transition' grant LERCO CZ.10.03.01/00/22_003/0000003 to V.Y. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the manuscript writing; or in the decision to publish the results.

Declaration of interests

The authors declare no competing interests.

Supplemental information

Supplemental information associated with this article can be found online at https://doi.org/10.1016/j.pt.2023.11.009.

[1]Life Science Research Centre, Faculty of Science, University of Ostrava, 710 00 Ostrava, Czechia

[2]Institute of Parasitology, Czech Academy of Sciences, 370 05 České Budějovice, Czechia
[3]Faculty of Sciences, University of South Bohemia, 370 05 České Budějovice, Czechia
[4]Faculty of Biology, M.V. Lomonosov Moscow State University, 119991 Moscow, Russia
[5]These authors contributed equally.

*Correspondence:
kostygov@gmail.com (A.Y. Kostygov) and
vyacheslav.yurchenko@osu.cz (V. Yurchenko).

https://doi.org/10.1016/j.pt.2023.11.009

References

1. Kostygov, A.Y. *et al.* (2021) Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open Biol.* 11, 200407
2. Maslov, D.A. *et al.* (2019) Recent advances in trypanosomatid research: genome organization, expression, metabolism, taxonomy and evolution. *Parasitology* 146, 1–27
3. El-Sayed, N.M. *et al.* (2005) Comparative genomics of trypanosomatid parasitic protozoa. *Science* 309, 404–409
4. Yurchenko, V. *et al.* (2021) Genomics of Trypanosomatidae: where we stand and what needs to be done? *Pathogens* 10, 1124
5. Bartholomeu, D.C. *et al.* (2021) Genomics and functional genomics in *Leishmania* and *Trypanosoma cruzi*: statuses, challenges and perspectives. *Mem. Inst. Oswaldo Cruz* 116, e200634
6. Jackson, A.P. *et al.* (2016) Kinetoplastid phylogenomics reveals the evolutionary innovations associated with the origins of parasitism. *Curr. Biol.* 26, 161–172
7. Alves, J.M.P. (2017) Amino acid biosynthesis in endosymbiont-harbouring Trypanosomatidae. In *The Handbook of Microbial Metabolism of Amino Acids* (D'Mello, J.P.F., ed.), pp. 371–383, CAB International
8. Billington, K. *et al.* (2023) Genome-wide subcellular protein map for the flagellate parasite *Trypanosoma brucei*. *Nat. Microbiol.* 8, 533–547
9. Gerasimov, E.S. *et al.* (2012) From cryptogene to gene? ND8 editing domain reduction in insect trypanosomatids. *Eur. J. Protistol.* 48, 185–193
10. Kostygov, A.Y. *et al.* (2020) *Vickermania* gen. nov., trypanosomatids that use two joined flagella to resist midgut peristaltic flow within the fly host. *BMC Biol.* 18, 187
11. Hamilton, P.B. *et al.* (2007) Patterns of co-evolution between trypanosomes and their hosts deduced from ribosomal RNA and protein-coding gene phylogenies. *Mol. Phylogenet. Evol.* 44, 15–25
12. Chmelová, Ľ. *et al.* (2021) Comparative analysis of three trypanosomatid catalases of different origin. *Antioxidants* 11, 46
13. Suh, S.O. *et al.* (2012) *Microbotryozyma collariae* gen. nov., sp. nov., a basidiomycetous yeast isolated from a plant bug *Collaria oleosa* (Miridae). *Antonie Van Leeuwenhoek* 102, 99–104
14. Opperdoes, F.R. *et al.* (2021) The remarkable metabolism of *Vickermania ingenoplastis*: genomic predictions. *Pathogens* 10, 68
15. Albanaz, A.T.S. *et al.* (2021) Genome analysis of *Endotrypanum* and *Porcisia* spp., closest phylogenetic relatives of *Leishmania*, highlights the role of amastins in shaping pathogenicity. *Genes* 12, 444