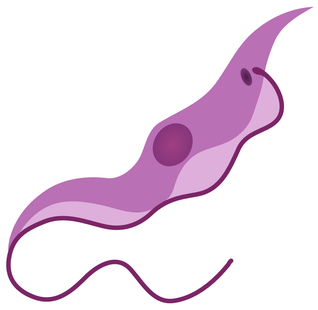

Phylogenetic framework to study evolution of traits in trypanosomatids

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

[1] Life Science Research Centre, Faculty of Science, University of Ostrava, Czechia; [2] Institute of Parasitology, Czech Academy of Sciences, České Budějovice, Czechia; [3]Faculty of Sciences, University of South Bohemia, České Budějovice, Czechia; [4]Faculty of Biology, Moscow State University, Moscow, Russia

Introduction

The family Trypanosomatidae is an extensively studied group of flagellates parasitizing vertebrates, arthropods, leeches, plants, and ciliates. Some members (*Leishmania* spp., *Trypanosoma brucei*, and *T. cruzi)* are important human pathogens. Trypanosomatids are categorized into monoxenous (with a single host) and dixenous (with two hosts, one of which is a vector). This family is characterized by many interesting biological features (polycistronic transcription, trans-splicing, RNA editing, glycosomes, etc). Some members have additional peculiarities such as unconventional genetic code, intracellular bacterial symbionts, or two flagella attached to each other. Recently the list of sequenced trypanosomatid genomes has significantly increased so that they are available for all genera of the family and for most subgenera of trypanosomes.

Methods

Using 44 publicly available genome assemblies of trypanosomatids, that of their closest outrgroup, the eubonid *Bodo saltans*, as well as three newly produced ones based on raw reads retrieved from GenBank, we created three phylogenetic datasets. These were: i) all available species of the subfamily *Leishmaniinae*, ii) all available species of *Trypanosoma,* and iii) all genera of Trypanosomatidae (except for *Leishmaniinae*). In each dataset, orthologous groups (OGs) of single-copy genes were identified and filtered using identity and length variation thresholds. Their sequences were aligned with MAFFT (L-INS-i algorithm) and trimmed with trimAl (strict method + 0.5 gap threshold). The trees were inferred in Phylobayes-MPI under GTR+CAT+G+I model with 20,046 – 58,848 generations to achieve convergence.

Distribution of catalases was estimated by searching the assemblies using a HMM profile created using reference trypanosomatid sequences of this protein. Maxicircle contigs were identified in genome assemblies using blastn and known sequences as queries. The identified contigs were manually annotated based on sequence homology to the features in reference data. Known full-length edited mRNA sequences were used to annotate unedited ORF fragments for partially edited genes and determine editing domain length. Relative boundaries of pan-edited cryptogenes were annotated manually with respect to robustly determined boundaries of non-edited or partially edited flanking genes.

Results and Discussion

Updated phylogeny of trypanosomatids

In general, the inferred trees agree with previous reconstructions based on a handful of genes, but there are 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 demonstrates that they cannot be separated into monophyletic taxa, thereby indicating the need for a taxonomic revision (Fig. 1A). We fully resolved the relationships between subgenera of *Trypanosoma* and confirmed the previously established basal split between aquatic and terrestrial species (Fig. 1B). We revealed that the latter group includes an early-diverging clade, in which African salivarian trypanosomes are sister to the subgenus *Squamatrypanum*, encompassing poorly studied parasites of squamates and small mammals (Fig. 1B). Importantly, the trypanosomes of Archosauria (crocodiles and birds) are not directly related, as exemplified by the relationships between the subgenera *Crocotrypanum* and *Trypanomorpha* (Fig. 1B).

In the whole-family tree, the previously suggested relatedness of *Wallacemonas* and *Sergeia* to Strigomonadinae is confirmed, but instead of being sister to each other, these genera consecutively branch off within the common clade, which is sister to Leishmaniinae (Fig. 1C). Another finding is the sister relationship between *Vickermania* and *Jaenimonas*, which could not be reliably inferred using single-gene analyses (Fig. 1C). In addition, our analysis revealed a relatively early branching of the subfamily Blastocrithidiinae, which diverged fourth after *Paratrypanosoma*, *Trypanosoma*, and *Blechomonas*.

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 (Fig. 2).

It has been previously demonstrated that catalase 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. 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 possesses this enzyme.

Another example is the distribution of RNA editing in 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. Our phylogenomic tree confirms this view: 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 (Fig. 2). As for other cryptogenes, the reduction of the edited domain likely occurred in parallel in different lineages, as evident from the comparison of *Kentomonas* and/or *Vickermania* with their respective relatives (Fig. 2). 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*.

The well-resolved phylogeny presented here for the first time encompasses almost all trypanosomatid genera and subgenera. We propose it as a framework to address a wide range of questions related to trait evolution in these amazing protists.

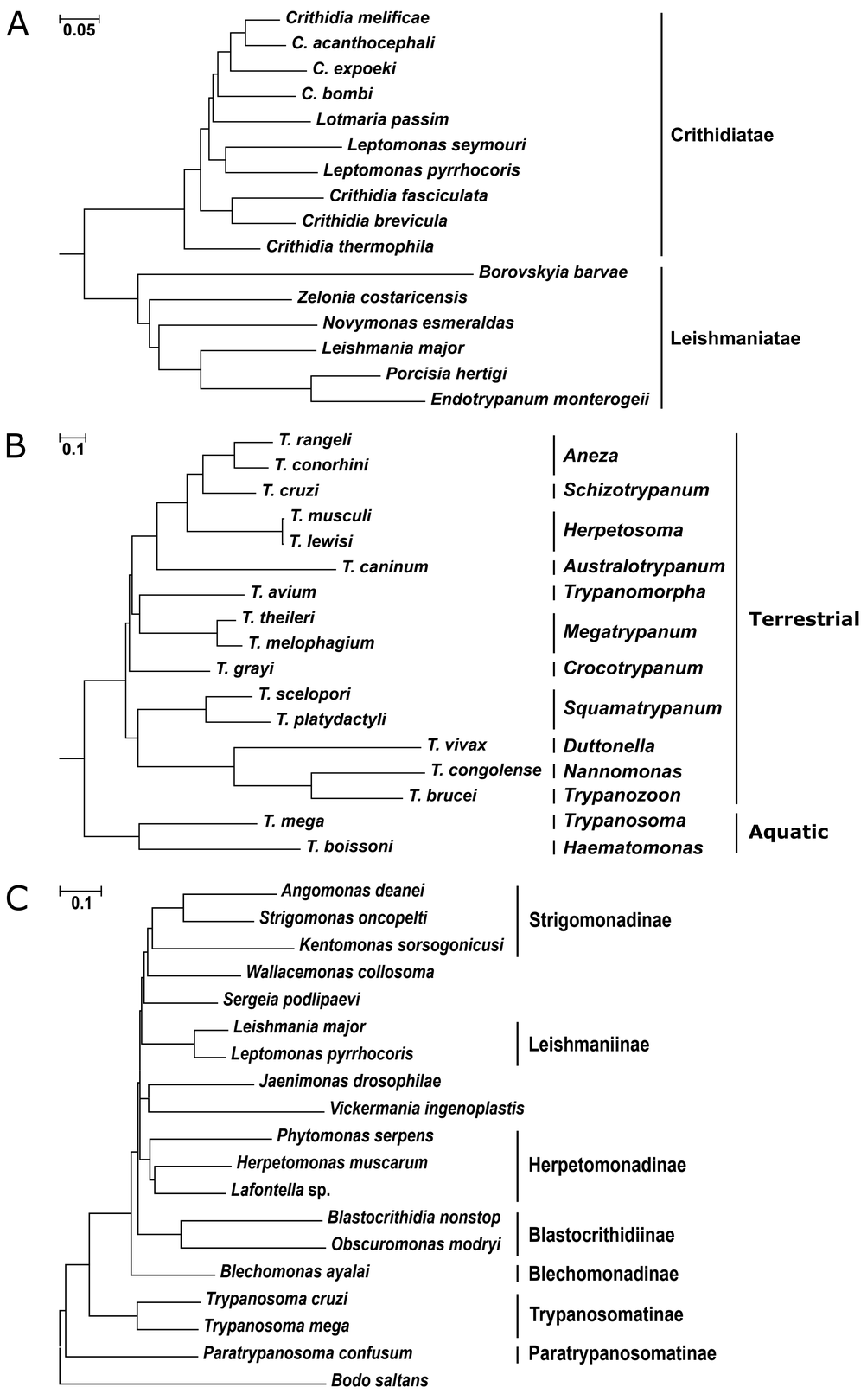

Fig. 1. Phylogenomic trees for the subfamily Leishmaniinae, the genus *Trypanosoma* (300 and 247 OGs, respectively, both at the species level) and the family Trypanosomatidae (except for Leishmaniinae) at the genus level (combination of two subdatasets of 124 and 255 OGs). All branches have posterior probability of 1.0 (A, B), or 0.99 – 1.0 (C). Scale bar corresponds to the number of substitutions per site. Outgroups are not shown in A an B.

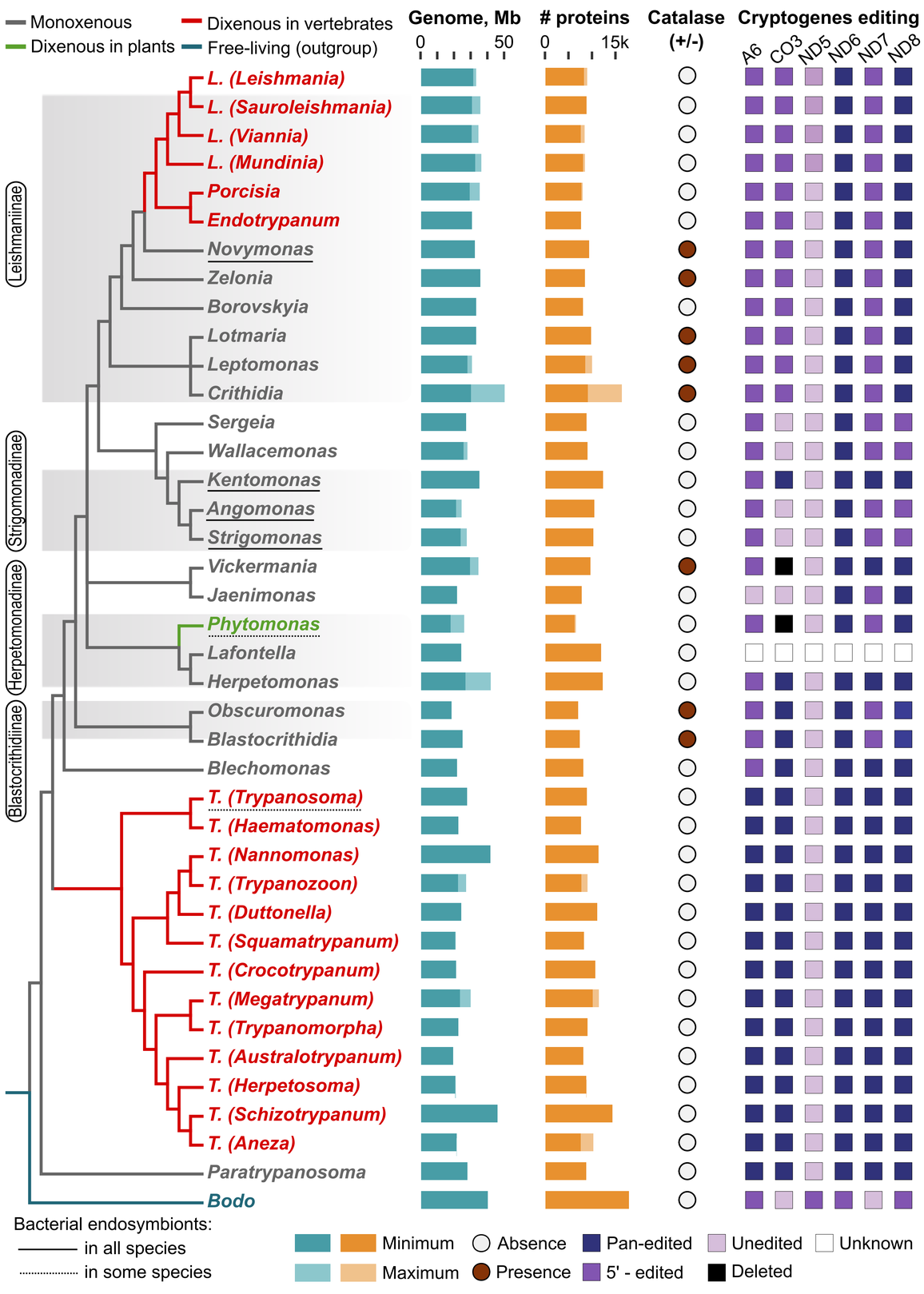

Fig. 2. Phylogenomic tree of Trypanosomatidae with mapped genomic features. It was combined from the trees reconstructed in this work (Fig. 1A-C) and a previously published one for the *Leishmania*/*Porcisia*/*Endotrypanum* clade (Albanaz et al., 2021). Lineages are colored according to their lifestyles. 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.

Co-funded by the European Union

Ministry of the Environment of the Czech Republic

LERCO