PHYLOGENOMICS OF FREE-LIVING NEOBODONIDS REVEALS THEY ARE A PARAPHYLETIC GROUP FROM WHICH ALL OTHER METAKINETOPLASTIDS ARE DESCENDED

A bioRχiv PREPRINT

Daryna Zavadska
https://orcid.org/0000-0002-8742-7044
zavadskadaryna@gmail.com
Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra),
Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain

Julia A. Packer
https://orcid.org/0009-0000-1969-6843
JPacker@dal.ca
Institute for Comparative Genomics, and Department of Biology, Dalhousie University
Halifax, Nova Scotia, Canada

Daria Tashyreva
https://orcid.org/0000-0002-1280-4095
tashyreva@gmail.com
Institute of Evolutionary Biology, Faculty of Biology, University of Warsaw
wirki i Wigury 101, 02-089, Warsaw, Poland

Ryoma Kamikawa
https://orcid.org/0000-0003-2648-8107
kamikawa.ryoma.7v@kyoto-u.ac.jp
Graduate School of Agriculture, Kyoto University
Kitashirakawa Oiwake-cho, Sakyo-ku, Kyoto 606-8502, Japan

Alastair Simpson
https://orcid.org/0000-0002-4133-1709
alastair.simpson@dal.ca
Institute for Comparative Genomics, and Department of Biology, Dalhousie University
6287 Alumni Crescent, PO Box 15000, Halifax, NS B3H 4R2, Canada

Daniel J. Richter
https://orcid.org/0000-0002-9238-5571
daniel.j.richter@gmail.com
Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra),
Passeig Marítim de la Barceloneta 37-49, 08003 Barcelona, Spain

April 9, 2026

ABSTRACT

Kinetoplastea is a major taxon of microbial eukaryotes that includes the well-known trypanosomatid parasites, species of which cause sleeping sickness, Chagas disease and leishmaniases in humans, as well as various animal and plant diseases. Free-living kinetoplastids are greatly understudied compared to their parasitic relatives, but are ecologically important microbivores, and collectively comprise the great majority of kinetoplastid diversity. For two decades kinetoplastids have been divided into prokinetoplastids and metakinetoplastids, with the latter further split into four orders, the most diverse of which is Neobodonida. However, the position of the root of the metakinetoplastids and whether neobodonids are a clade has remained unclear due to a lack of multi-gene data from free-living kinetoplastids, particularly neobodonids. Here, we present transcriptomic data for eleven newly or recently cultivated free-living neobodonid kinetoplastids. Phylogenomic analyses of a *de novo* generated data set of 444 inferred orthologs and 49 taxa robustly resolve the kinetoplastid tree, including the position of the root of metakinetoplastids. This divides metakinetoplastids such that Trypanosomatida, Eubodonida, Parabodonida, Allobodonidae (formerly neobodonid clade 1E) and neobodonid clade 1D fall on one side, while neobodonid clades Nd6, 1B (Rhynchomonadidae) and 1C fall on the other. Neobodonids are thus inferred to be a paraphyletic group from which all other metakinetoplastids descend. This analysis is the most thorough examination of metakinetoplastid phylogeny thus far, and forms a new basis for tracing the evolutionary history of the entire kinetoplastid group.

Keywords Phylogenomics · Kinetoplastid · Metakinetoplastid · transcriptome

1 Introduction

The kinetoplastids (Kinetoplastea; Euglenozoa) are a major group of heterotrophic protists that are of substantial medical and ecological importance, and that have bizarre basic cell biology Gibson (2017). Almost all kinetoplastids are uni- or bi-flagellate unicells that are either free-living phagotrophs, or parasites/commensals of other eukaryotes (mainly animals) Gibson (2017); Goodwin et al. (2018). Their principal characteristic feature is the kinetoplast, a large aggregation (or aggregations) of DNA within the mitochondrion that typically includes two distinct categories of molecules, minicircles and maxicircles Lukeš et al. (2002). The minicircles, principally, encode small guide RNAs that permit editing of primary transcripts of the protein-coding genes that are encoded on maxicircles. Each edit inserts or deletes a fixed number of uracils at a specific position in the transcript, as specified by the guide RNA sequence. The editing can be extremely extensive, with dozens of different guide RNAs required for all the edits needed to correctly synthesise the small complement of mitochondrially-encoded proteins Lukeš et al. (2002). Other characteristic features of most kinetoplastids include paraxonemal rods within the flagella, a feeding apparatus supported longitudinally by microtubules derived from a particular flagellar root, addition of spliced leaders to nuclear transcripts, and much of glycolysis being localised to a modified peroxisome (the glycosome) Brugerolle et al. (1979); Simpson (1997); Simpson et al. (2006); Tashyreva et al. (2022). These four latter features, however, can be found in some other Euglenozoa: diplonemids, euglenids and/or symbiontids Simpson (1997); Butenko et al. (2020).

The best studied kinetoplastids are the trypanosomatids (Trypanosomatida). This diverse clade of uniflagellate parasites includes the agents causing sleeping sickness (*Trypanosoma brucei*), Chagas disease (*Trypanosoma cruzi*) and leishmaniases in humans, nagana in livestock, and several plant diseases Steverding (2008); Hide (1999). Numerous studies have examined many aspects of trypanosomatid cell biology, biochemistry and molecular biology, including most of the research on the kinetoplast and mitochondrial RNA editing described above, but also immune system evasion by Variant Surface Glycoprotein (VSG) switching in *T. brucei*, amongst many other phenomena Lukeš et al. (2023). All other kinetoplastids are historically known as bodonids. These are almost always motile biflagellates, and range from free-living phagotrophs to (probable) necrotrophs to symbionts/commensals or parasites of (mostly aquatic) animals Gibson (2017); Goodwin et al. (2018). It was generally supposed that bodonids were paraphyletic; this was confirmed by molecular phylogenetics, which two decades ago was used as the basis of the subdivision of kinetoplastids/bodonids into the major taxa still used today Moreira et al. (2004); Kostygov et al. (2021); Adl et al. (2019). Analyses of various sequence data consistently divide known kinetoplastids into Prokinetoplastina/Prokinetoplastia and Metakinetoplastina/Metakinetoplastia Callahan et al. (2002); Simpson et al. (2004); Moreira et al. (2004); Cenci et al. (2016); Tikhonenkov et al. (2021). Prokinetoplastina is now known to contain free-living phagotrophs Tikhonenkov et al. (2021) in addition to the fish parasite *Ichthyobodo* and the highly reduced parasomes (e.g. *Perkinsela*), which are vertically transmitted endosymbionts of paramoebid amoebae Kostygov et al. (2021). Metakinetoplastina includes Trypanosomatida plus three additional major taxa: Eubodonida, Parabodonida and Neobodonida Moreira et al. (2004); Kostygov et al. (2021). Eubodonida currently contains free-living biflagellates that may all be referable to the morphospecies *Bodo saltans*. Parabodonida includes the free-living *Parabodo* and *Procryptobia* plus the commensalic/parasitic *Cryptobia* and *Trypanoplasma*. Neobodonida is much more diverse, including a dozen nominal genera across the free-living to parasite spectrum; examples include *Rhychomonas*, *Dimastigella*, *Rhynchobodo*, the parasitic *Azumiobodo*, and the clearly non-monophyletic *Neobodo* itself. Phylogenies of 18S rDNA sequences in turn divide neobodonids *sensu lato* into several subclades; 1A, 1B (Rhynchomonadidae), 1C, Nd6 (sometimes considered part of 1C), 1D, and 1E (Allobodonidae) von der Heyden et al. (2004); von der Heyden and Cavalier-Smith (2005); Goodwin et al. (2018).

Attempts to understand the evolutionary history of kinetoplastids, such as the origins of parasitic kinetoplastids from free-living ancestors, or the origins of the various unusual features of kinetoplastid cell/molecular biology, require a well resolved and accurate backbone phylogeny of the group. To date, phylogenies of 18S rDNA sequences and 1-2 heat shock proteins have not resolved the deep branches within Metakinetoplastina. Unrooted analyses of Metakinetoplastina most often place eubodonids adjacent to trypanosomatids, though generally with poor support Simpson et al. (2002); Goodwin et al. (2018). Rooted analyses, with prokinetoplastids and/or other outgroups included, sometimes (in the case of 18S rDNA) or usually (in the case of heat shock proteins) place some or all neobodonids as the deepest branch within metakinetoplastids Simpson et al. (2002, 2004); Goodwin et al. (2018), indicating that eubodonids, then parabodonids, are successive sisters to trypanosomatids Figure 1-A. Others place the root elsewhere, including within trypanosomatids, or with trypanosomatids as a whole sister to all other metakinetoplastids von der Heyden et al. (2004); von der Heyden and Cavalier-Smith (2005); Moreira et al. (2004) Figure 1-C. While 18S rDNA

datasets have by far the best taxon sampling, an extremely long basal branch separates metakinetoplastids from any outgroup Simpson et al. (2002); Goodwin et al. (2018); Tikhonenkov et al. (2021). Predictably, the root position is highly unstable and any particular position is generally very poorly supported Goodwin et al. (2018). This pattern is less extreme in protein datasets, however taxon sampling is much poorer Simpson et al. (2004). Notably, no protein dataset includes all of the neobodonid subgroups, with the complete absence of Allobodonidae (1E) being especially problematic. Allobodonidae is robustly recovered as the deepest neobodonid branch in unrooted 18S rDNA trees, in other words, all other neobodonids form a robust clan von der Heyden et al. (2004); Goodwin et al. (2018); Packer et al. (2025). This means that prior protein analyses simply cannot distinguish between a rooting position next to neobodonids (Neobodonida monophyly) Figure 1-A or within neobodonids (neobodonid paraphyly, Figure 1-B,D).

This problem of underpowered taxon sampling is even more acute in the few phylogenomic analyses that are specifically designed to infer the deep-level phylogeny of kinetoplastids. Studies to date strongly support the general neobodonids deep scenario (and the sister relationship of eubodonids and trypanosomatids) Figure 1-A,B, but the sampling of bodonids is extremely limited, and in particular they include at most two neobodonids, from Rhynchomonadidae and 1C Deschamps et al. (2011); Yazaki et al. (2017).

Resolving the deep-level relationships within Metakinetoplastina is crucial for a rational high-level systematics of the group and to further resolve the evolution of the multiple enigmatic, bizarre and unique traits found in kinetoplastids, especially the medically important trypanosomatids. Here, we report a phylogenomic analysis of kinetoplastids with markedly enhanced taxon sampling of free-living metakinetoplastids, enabled by new, or recently reported Packer et al. (2025) cultivation efforts. This improves representation across the known diversity of neobodonids in particular, and notably includes allobodonids for the first time, as well as 1D and Nd6. The examined data is a *de novo* curation of 444 inferred single-copy (or nearly single-copy) orthogroups present in a high proportion of available kinetoplastid genomes/transcriptomes. Our analyses strongly support a rooting of metakinetoplastids within neobodonids, suggesting a neobodonid-like ancestor for all metakinetoplastids, and indicating that the high-level (ordinal) systematics of the group will need revision in the future.

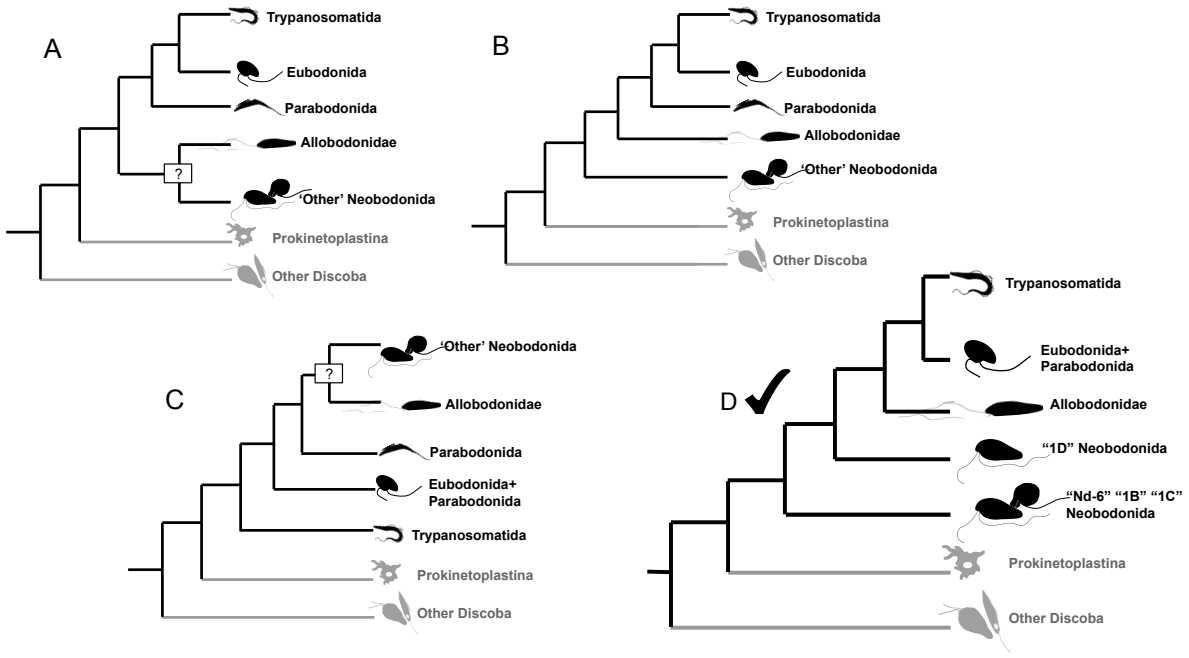

Figure Figure 1: **Different scenarios of Metakinetoplastina phylogeny rooting and its effect on the topology of the tree.** A-C - various hypotheses on the rooted topology of the phylogeny of the Metakinetoplastina phylogenetic tree outlined in the introduction. D - the topology inferred in this study. "?" indicates an uncertain node in one of the hypothetical topologies.

2 Materials and Methods

2.1 Sampling, cultivation and light microscopy

Isolates BEAP0154-"D44", BEAP0117-"13G" and BEAP0071-"10D" were obtained from plankton samples of subsurface water near Blanes Bay, Mediterranean Sea, Spain (GPS coordinates 41.66667 N, 2.8 E) collected during the year 2022. Isolates BEAP0230-"RhMon" and BEAP0214-"G65133" were obtained from subsurface water of the Caribbean Sea, roughly 10 km south of the island of Grenada (GPS coordinates 11.893732 N, 61.699506 W), in February 2023. We enriched the plankton samples with the nutrient component of RS medium (dilutions from 1:5 to 1:500), and obtained stable mixed protist cultures from the enrichment samples, which were cultivated in RS medium (`https://mediadive.dsmz.de/medium/P4`, `https://mediadive.dsmz.de/medium/P5`) with nutrient (NM) to non-nutrient (NNM) component ratios between 1:50 and 1:500, at 17°C (for BEAP0154-"D44", BEAP0117-"13G" and BEAP0071-"10D") and at 23°C (BEAP0230-"RhMon" and BEAP0214-"G65133"). The isolates we used in these studies were monoclonal cultures obtained from initial stable mixed cultures by performing 3 to 5 sequential rounds of dilution to extinction. Isolates BEAP0117-"13G", BEAP0154-"D44", BEAP0230-"RhMon" and BEAP0214-"G65133" were all initially and routinely cultivated with a 12h:12h light:dark cycle, while isolate BEAP0071-"10D" was cultivated in darkness. We note that, if necessary, all of them are also capable of growth in darkness.

Isolate "OB23" was established from a seawater sample collected at a depth of 5 m at the entrance of Osaka Bay (GPS coordinates 34.324444 N, 135.120833 E), Japan, in July 2023. The sample was enriched in Daigos IMK medium (Fujifilm Wako Pure Chemical Corporation). Single cells were then isolated from the enrichment culture using a glass pipette, followed by several sequential rounds of serial dilution to extinction. The established culture was maintained at 20°C under dark conditions in Daigos IMK medium.

Freshwater isolates "BAB10-KIN1", "INU10-KIN1", and "KUC3-KIN1" were isolated from water samples collected in October 2023 in the Masurian Lakeland, Poland. "BAB10-KIN1" originated from a sample of the hypolimnion of Lake Babity (53.719288 N, 21.129903 E), while "INU10-KIN1" and "KUC3-KIN1" were isolated from samples of the epilimnion of Lake Inulec (53.805704 N, 21.491791 E) and Lake Kuc (53.815844 N, 21.398909 E), respectively. To enrich for kinetoplastids, 10 mL of each sample were incubated with an autoclaved rice grain for 7 days. Individual kinetoplastid cells were manually picked, together with associated bacteria, using a glass microcapillary and transferred to separate wells of 12-well microplates containing NM component diluted 1:100 in Bolds Basal Medium with threefold nitrogen and vitamins (3N-BBM Bischoff (1963); `https://mediadive.dsmz.de/medium/C15`). The isolates were routinely maintained in 3N-BBM supplemented with 1:100 dilution of NM at +18°C in the dark.

The isolation of *Novijibodo darinka*-"Gemkin" and *Allobodo yubaba* BEAP0186 "16Ckin" was described previously, in Packer et al. (2025).

All isolates in this study were grown as xenic cultures (i.e., with several uncharacterized bacterial species).

Light microscopy data from selected isolates used in this study (BEAP0230-"RhMon", BEAP0214-"G65133", BEAP0154-"D44", BEAP0117-"13G", BEAP0071-"10D", "BAB10-KIN1", "INU10-KIN1" and "KUC3-KIN1") was obtained with a Zeiss Axio Observer.Z1 inverted microscope with a Differential Interference Contrast (DIC) Plan-Apochromat 63x Oil immersion objective (NA=1.4). The imaging setup was equipped with a HDCamC13440-20CU Hamamatsu camera coupled with ZEN image acquisition software. Imaging data was further processed in Fiji Schindelin et al. (2012). Transmission Electron Microscopy was performed at Electron Cryomicroscopy Unit at the Scientific and Technical Centers of the University of Barcelona. Cells were collected by centrifugation (2000 x *g* for 5 min), resuspended in 20% BSA in artificial sea water and cryo-immobilized using a Leica HPM100 high-pressure freezer (Leica Microsystems, Vienna, Austria). Samples were freeze-substituted with 2% osmium tetroxide and 0.1% uranyl acetate dissolved in the pure acetone at -90°C for 72 hours in an EM AFS2 (Leica Microsystems, Vienna, Austria), followed by gradual warming up to RT and infiltration of Epon-812 resin and polymerization in Epon-812 at 60°C for 48 hours. Ultrathin Sections of 60 nm were obtained with a UC6 ultramicrotome (Leica Microsystems, Vienna, Austria) and placed on Formvar-coated copper grids. Sample Sections were stained with 2% uranyl acetate for 30 min, lead citrate for 5 min and examined in a TEM Jeol JEM 1010 (Gatan, Japan) equipped with a tungsten cathode. Images were acquired at 80 kV with a 1k x 1k CCD Megaview III camera. Complete imaging data, including movies, are available via Figshare at the link `https://doi.org/10.6084/m9.figshare.30214630`.

2.2 RNA isolation and transcriptome sequencing

The *Novijibodo darinka* isolate Gemkin was grown to high abundance in a 400 ml parafilm-sealed vented culture flask, initially containing 20 ml of cell culture, 100 ml of CR10 media (Gigeroff et al. (2023), with CR number indicating salinity in ppt), and 12 autoclave-sterilised wheat grains. The culture was grown for 12 days in a 21°C dark incubator before being pelleted via centrifugation (1500 x *g* for 12 min, then 6500 x *g* for 3 min). The pelleted cells were lysed using 4 ml TRIzol™ (ThermoFisher), and stored at -20°C for 2 weeks. Total RNA was isolated using the standard TRIzol™ protocol, except that the RNA was precipitated at -20°C overnight. Total RNA was washed twice using the RNeasy PowerClean Pro cleanup kit (QIAGEN). The RNA was stored at -80°C before being sent to Génome Québec for poly-A selection, cDNA library preparation, and paired-end sequencing (100 bp) using Illumina NovaSeq.

Isolate OB23 was cultivated as described above in eight 50-mL flasks (Violamo) containing Daigos IMK medium. After two weeks of cultivation, the cells were pelleted by centrifugation at 2,000 x *g* for 10 min. RNA was extracted from the pelleted cells using TRIzol reagent and Phasemaker Tubes (Thermo Fisher Scientific) according to the manufacturer's instructions. Total RNA was then used for library construction using the MGIEasy Fast RNA Library Prep Set (MGI Tech). The resulting libraries were circularized using the MGIEasy Dual Barcode Circularization Kit (MGI Tech) and converted into DNA nanoballs (DNBs) using the DNBSEQ-G400RS High-throughput Sequencing Kit and High-throughput Pair-End Sequencing Primer Kit (App-D) (MGI Tech). Finally, 150-bp paired-end sequencing was performed on a DNBSEQ-G400 platform (MGI Tech).

Total RNA from *Allobodo yubaba* (isolate BEAP0186-"16Ckin"), and from the remaining isolates used in this study (BEAP0230-"RhMon", BEAP0214-"G65133", BEAP0154-"D44", BEAP0117-"13G", BEAP0071-"10D" as well as "BAB10-KIN1", "INU10-KIN1" and "KUC3-KIN1") was extracted following the procedure from Richter et al. (2018). All of these isolates used in transcriptomics studies were strictly monoclonal (i.e., they were initiated from a single eukaryotic cell and any co-isolated environmental bacteria), except for "INU10-KIN1", for which the culture originated from a few manually picked cells . For RNA isolation, cultures were grown in RS medium (`https://mediadive.dsmz.de/medium/P4`, `https://mediadive.dsmz.de/medium/P5`), of nutrient (NM) to non-nutrient (NNM) component ratios between 1:20 and 1:500 for marine isolates (BEAP0230-"RhMon", BEAP0214-"G65133", BEAP0154-"D44", BEAP0117-"13G", BEAP0071-"10D") or 1:100 dilutions of NM component prepared in distilled water in 3N-BBM (Bold's Basal Medium) Bischoff (1963) for freshwater isolates ("BAB10-KIN1", "INU10-KIN1" and "KUC3-KIN1"), in 75 cm^2 rectangular canted-neck cell culture flasks with vented caps (353136, Corning Life Sciences) and pelleted by centrifugation (13000 x *g* for 20 min, at +4°C). RNA isolation was performed with an RNAqueous kit (AM1912, ThermoFisher Scientific), followed by genomic DNA digestion using a TURBO DNA-free kit (AM1907, ThermoFisher Scientific). The kits were used according to the manufacturer's protocols with the modifications described in Richter et al. (2018). Double the normal rounds of polyA selection were applied to the total RNA before cDNA synthesis (to reduce the amount of prokaryotic RNA in the sample, which is common in non-axenic cultures). The integrity of total RNA was checked using Bioanalyzer 2100 RNA Pico chips (Agilent Technologies) at the CRG Genomics Core Facility in Barcelona, and total RNA was stored at -80°C before being sent to the same facility for poly-A selection, library preparation, and sequencing. Stranded libraries were prepared using the NEBNext Ultra II Directional RNA Library Prep Kit. Finally, the RNA was sent for paired-end sequencing (150 bp) on an Illumina NextSeq 2000.

2.3 Transcriptome assembly, decontamination, protein prediction and redundancy reduction

The procedure described below and in Figure S3 was performed for all the transcriptomes obtained in this study (isolates "Gemkin", "OB23", BEAP0186-"16Ckin", BEAP0230-"RhMon", BEAP0214-"G65133", BEAP0154-"D44", BEAP0117-"13G", BEAP0071-"10D", "BAB10-KIN1", "INU10-KIN1" and "KUC3-KIN1").

2.3.1 Read trimming and transcriptome assembly

Raw reads were trimmed using fastp(v0.20.1) Chen et al. (2018); depending on the sequencing technology used, settings varied (see exact command list at `https://github.com/beaplab/kineto_phylo/blob/main/MASTER_commands_V2.sh`). Assembly was performed using RNAspades v3.13.1 Bushmanova et al. (2019), in stranded mode (with the exception of "OB23", which was assembled in non-stranded mode), with the other settings left as their defaults.

2.3.2 Identification of putative contaminants and creating the database of contaminant genomes

Potential contaminants were primarily identified by searching for 18S ribosomal RNA gene sequences of the non-target eukaryote species and for prokaryotic 16S ribosomal sequences in the assemblies.

18S and 16S of contaminants were retrieved by blasting 18S and 16S sequences of *Mus musculus* (NR_003278) and *Escherichia coli* (NR_114042) against the assembled contigs. Matches were manually filtered out, and target contigs producing reasonable match length, E-value and % identity were selected (see table on FigShare at `https://doi.org/10.6084/m9.figshare.30214630` for the exact list of contigs and contaminants). These contigs that putatively contained 18S or 16S of contaminants were used as queries to perform blastn against the NCBI nt database (a custom Perl script was used as a wrapper for performing blastn of multiple queries vs. multiple targets, available on Github `https://github.com/beaplab/kineto_phylo/blob/main/Suppl_v1_scripts/web_blast.pl`). A single top hit was selected for each contig. NCBI taxonomy IDs of the targets from the top hits were retrieved.

In addition to the taxonomy IDs matching 16S- and 18S-containing contigs identified by the BLAST approach, we also retrieved taxonomy IDs of 18S sequences assembled by phyloFlash v3.3b3 Gruber-Vodicka et al. (2020) from the raw transcriptomic reads ("Full-length SSU rRNA sequences assembled by SPAdes, matched to SILVA database with Vsearch" list of PhyloFlash output).

Genomes corresponding to all the taxonomy IDs of potential contaminants were downloaded using the Entrez v14.6 and NCBI datasets v16.22.1 tools. If there were multiple genomes available for a single taxonomy ID, up to 5 genomes were selected for download, and RefSeq genomes were given preference.

2.3.3 Removal of contamination

For decontamination, assembled contigs were used as blastn queries against the database of putative contaminant genomes obtained in the previous step (a custom Perl script was used as a wrapper for performing blastn of multiple queries vs. multiple targets; it is available on Github `https://github.com/beaplab/kineto_phylo/blob/main/Suppl_v1_scripts/run_BLAST.pl`).

The distribution of % identity, E-value and match length was visualised for blast hits of 60,000 randomly selected contigs from all assemblies vs. putative contaminant genomes Figure S5 using a custom R script available on Github `https://github.com/beaplab/kineto_phylo/blob/main/Suppl_v1_scripts/blast_res_viz.Rmd`.

The minimum threshold for blastn hit match length was set to 150 for all assembly-contaminant genome pairs, based on visual inspection of the aforementioned distribution. Minimum thresholds of % identity were selected individually for each pair of assembly and contaminant genome, based on an automated algorithm designed to select a value separating identical or nearly identical matches resulting from contamination versus true matches resulting from sequence homology (see the same custom R script for details).

Contigs that matched contaminant genomes with % identity higher than the automatically identified cutoff and with the length of match exceeding 150 bp, were removed from the corresponding assemblies (using seqkit v0.10.0 Shen et al. (2016)).

2.3.4 Removal of Cross-contamination

To remove potential cross-contamination among transcriptomes which were sequenced together on the same flow cell (BEAP0230-"RhMon"+BEAP0214-"G65133", BEAP0186-"16Ckin"+BEAP0154-"D44"+BEAP0117-"13G"+BEAP0071-"10D" and "BAB10-KIN1"+"INU10-KIN1"+"KUC3-KIN1" were sequenced together), the WinstonCleaner `https://github.com/kolecko007/WinstonCleaner` tool was run with default parameter values. WinstonCleaner removes cross-contamination by mapping trimmed reads from the species sequenced together against the assembled contigs for each species in order to estimate coverage, and when a pair of contigs from different assemblies is found to have a high sequence similarity, the contig with lower read coverage is considered a contaminant and is removed from the corresponding assembly. The cutoffs of minimum % identity for the overlapping contigs to be considered cross-contaminants were automatically identified for each assembly pair and additionally verified by manually inspecting the distribution of % identity of all contigs in assembly pairs.

2.3.5 Protein prediction and assembly completeness

Protein prediction was done with the TransDecoder version v5.0.1 tool run in stranded mode (with the exception of "Gemkin" and "OB23", for which it was run in non-stranded mode) `https://github.com/TransDecoder/TransDecoder` to produce predicted CDSs and proteins for each decontaminated assembly.

2.4 Obtaining and processing data from other species

Data for other species used in the phylogenomic inference was retrieved from various sources (see table on FigShare at `https://doi.org/10.6084/m9.figshare.30214630`). Raw reads for *Cruzella marina* (ERR13669942) and *Azumiobodo hoyamushi* (SRR10586159) were downloaded from NCBI and assembled with RNAspades, in stranded mode, with other settings default. Assembled contigs of *Neobodo designis* and *Trypanoplasma/Cryptobia borreli* were obtained from EukProt Richter et al. (2022). Assemblies of these four kinetoplastid species (*Neobodo designis, Trypanoplasma/Cryptobia borreli, Cruzella marina, Azumiobodo hoyamushi*) were cleaned using the procedure to remove contamination described above. Assembled contigs of the outgroup species *Pharyngomonas kirbyi* were also obtained from EukProt Richter et al. (2022).

For the data from *Tsukubamonas globosa* and *Ploeotia vitrea*, the same procedures were applied as for *Pharyngomonas kirbyi*. However, these two species were excluded from the final set of taxa used in phylogenomic inference, because of low assembly completeness.

Proteins were predicted as described above for *Cruzella marina, Azumiobodo hoyamushi, Neobodo designis, Trypanoplasma/Cryptobia borreli* and *Pharyngomonas kirbyi* (further details are provided in S2.2).

Proteomes/CDS from the rest of the species were downloaded from the sources listed in table on FigShare at `https://doi.org/10.6084/m9.figshare.30214630`, CDSs were translated to amino acid sequences, if necessary, and used in the subsequent procedures in an unaltered state (without additional decontamination or quality control) (see S2.2).

2.5 Redundancy reduction

A redundancy-reduced proteome set was generated from the initial set of proteomes using CD-HIT v4.8.1 Li and Godzik (2006); Fu et al. (2012). The redundancy-reduced set ("NR90") retained proteins with <90% amino acid global sequence identity for each proteome (the default identity metric is calculated by CD-HIT as the number of aligned identical amino acids divided by the full length of the shorter sequence Li and Godzik (2006); Fu et al. (2012)).

2.6 Orthology inference and verification

The final dataset used for orthology inference contained 11 new predicted proteomes produced in this study, and 38 more predicted proteomes: 28 from kinetoplastid species and 10 from other Discoba lineages (as outgroups), resulting in 49 proteomes in total.

Orthogroups (OGs) were inferred using OrthoFinder v2.5.5 Emms and Kelly (2019), using the "diamond_ultra_sens" option for the sequence search algorithm and the "msa" option for gene tree inference, with other options left as their defaults.

Initially, 72407 OGs were identified in the set of 49 proteomes. The subset of OGs used for single-OG tree reconstruction and further manual curation was selected based on the species represented and on the number of proteins per species in each OG. Within each OG, we estimated the mean and the standard deviation of the number of proteins present per species, and we counted the number of species with at least one protein. Based on the distributions of these three parameters across all 72407 OGs (see Figure S8), we identified a set of thresholds which would theoretically allow us to produce a subset OGs containing the highest number of single-copy orthologs (i.e., one copy per species) in the largest number of taxa. The summary statistics on the OGs were calculated using a custom R script available on Github `https://github.com/beaplab/kineto_phylo/blob/main/Suppl_v1_scripts/Orthofinder_stats_UNIVERSAL_V2.Rmd`.

Our thresholds resulted in the selection of OGs meeting all three of the following criteria: (i) mean number of proteins per single species between 0.9 and 1, inclusive, (ii) standard deviation of the number of proteins per single species less than or equal to 0.5, and (iii) number of species with at least one protein present greater than 35 (out of 49 species in total).

The subset obtained by applying these criteria contained 461 OGs. In order to produce a set of single-copy orthologous sequences, first, proteins within each OG were aligned with MAFFT v7.490 Katoh and Standley (2013), using the L-INS-i ("–localpair") strategy, with the iteration limit set to 16. Resulting alignments were trimmed using trimAl v1.4.rev22 Capella-Gutiérrez et al. (2009) with the "–gappyout" option, and maximum likelihood single-OG trees were inferred using IQ-TREE v1.6.12 Nguyen et al. (2015) with node supports estimated by 1000 ultrafast bootstrap replicates, and other options as their defaults.

Each single-OG tree was manually examined to mark inparalogs, contaminant sequences, and other artefacts for removal from the corresponding OG. As a result of single-gene tree checks, 17 OGs were completely removed from further analysis, 377 OGs were edited by removing some of the proteins, and 67 OGs were left unedited. Protein sequences from the set of 461 OGs, alignments, single-OG trees, as well as the lists of sequences and OGs that were removed by manual inspection, can be found at table S3 and orthogroup directory on FigShare at `https://doi.org/10.6084/m9.figshare.30214630`.

2.7 Ortholog alignment and trimming

After paralog and artefact removal, proteins of each OG were re-aligned with MAFFT v7.490 Katoh and Standley (2013), using the L-INS-i ("–localpair") strategy, with the iteration limit set to 1000. Resulting alignments were trimmed using trimAl v1.4.rev22 Capella-Gutiérrez et al. (2009) with the "–gappyout" option. Trimmed alignments of all OGs were concatenated using PhyKit v2.0.1 Steenwyk et al. (2021).

2.8 Phylogenomic tree inference

Bayesian tree inference was performed with PhyloBayes-mpi v1.9 Lartillot et al. (2013). Five independent chains were run with the CAT+GTR+G4 model, sampling once every generation. The run was stopped at the 16000th generation, at which point the five chains did not converge, however the only differences between them were in the inter-relationships among the three prokinetoplastids. A consensus tree topology was obtained with a burn-in of 4000 generations (

25% of the total number of generations per chain). The mean discrepancy across all bipartitions of all three chains was 0.00631579.

A maximum likelihood analysis was run using IQ-TREE v2.3.6 Minh et al. (2020). The LG4M+R10 model was determined by the IQ-TREE ModelFinder Kalyaanamoorthy et al. (2017) as the best-fitting, according to BIC, as well as AIC and AICc. The robustness of the most likely ML tree with the LG4M+R10 model was inferred using ultrafast bootstrapping (1000 replicates) Hoang et al. (2018) along with standard bootstrapping (200 replicates).

Trees were visualised in tvBOT Xie et al. (2023) and edited for presentation with Inkscape.

3 Results

We sequenced and *de novo* assembled the transcriptomes of 11 isolates of neobodonids. Two of these, *Novijibodo darinka* and *Allobodo yubaba*, were described previously Packer et al. (2025), whereas the remaining 9 isolates remain undescribed Figure S1, Figure S2. We combined the predicted proteins of these 11 new transcriptomes with publicly available genomes and transcriptomes representing Metakinetoplastina, Prokinetoplastina and discoban ougroups and performed *de novo* orthogroup inference. We selected single-copy (or nearly single-copy) orthogroups, aligned them individually and concatenated the alignments. The final phylogenomic matrix used in this study contained 144,376 sites from 444 *de novo* inferred OGs and 49 species, with an average OG occupancy of 90%, and an average of 86.9% of sites present per species (i.e. non-gapped, non-missing, non-unknown) Figure S9.

Maximum Likelihood and Bayesian analyses recovered identical phylogenies with maximum support at almost all nodes Figure 2; the exceptions were the two internal nodes within Prokinetoplastina, which were not well resolved in

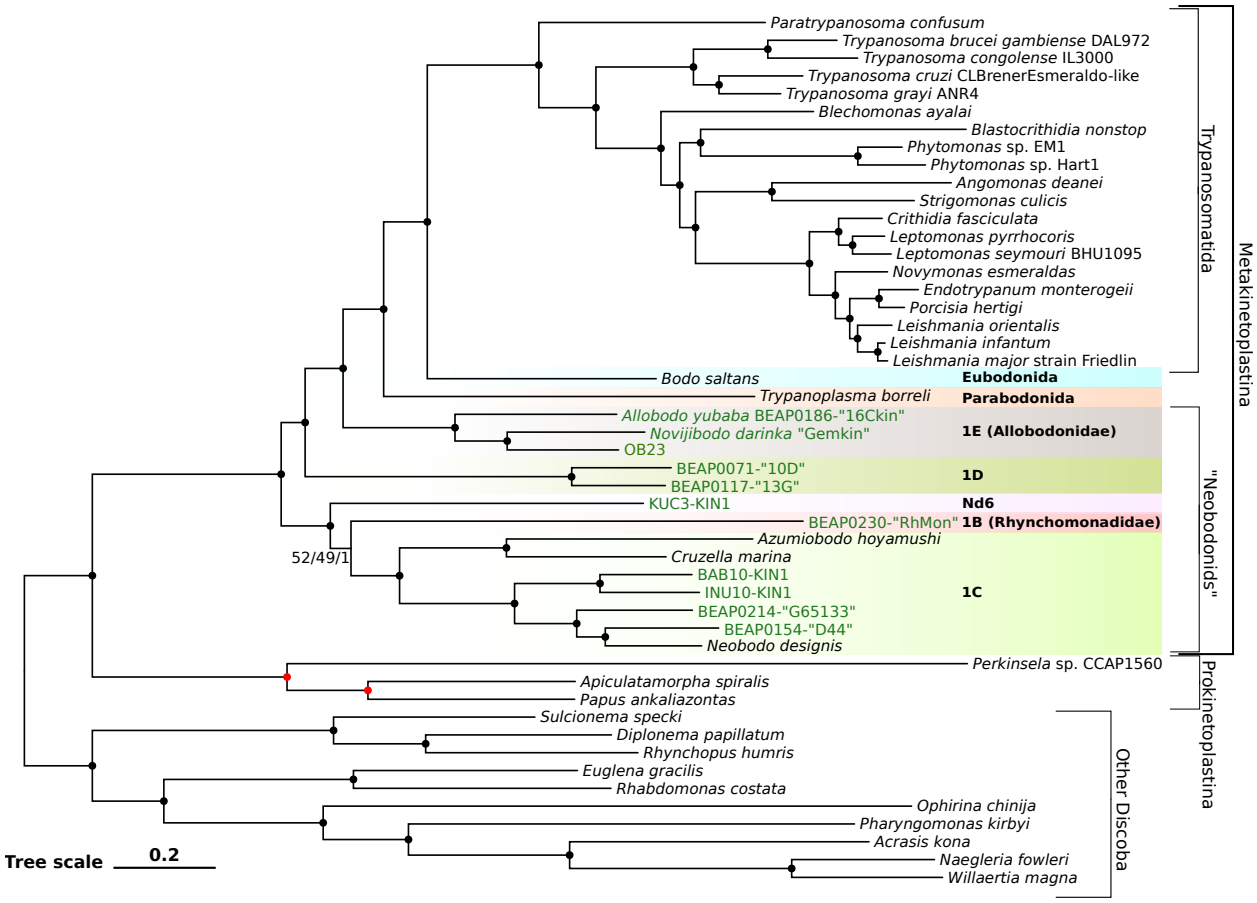

Figure Figure 2: **444-protein phylogeny of Kinetoplastea.** The topologies of the Maximum Likelihood (ML) and Bayesian phylogenies are identical (apart from the internal nodes of Prokinetoplastina, highlighted in red, which are different in the Bayesian phylogeny, see Figure S10). All nodes indicated by black circles are recovered in both ML and Bayesian analyses and have 100% support from 1000 ultrafast bootstrap replicates, as well as 200 full bootstrap replicates, and a posterior probability of 1. Nodes without full bootstrap support have ultrafast bootstrap, full bootstrap and posterior probability indicated and separated by slashes. The scale represents expected substitutions per site; branch lengths shown are those of the Maximum ML phylogeny inferred under the LG4M+R10 model. isolates for which the data were produced in this study are highlighted in green. Tree is shown rooted arbitrarily to maximise visual resolution within kinetoplastids.

the Bayesian analysis, while the node splitting 1C neobodonids and Rhynchomonadidae had poor support in the ML analysis Figure 2.

Monophyly is recovered for each major lineage included as an outgroup: Diplonemida, Euglenida and Heterolobosea (Jakobida was also included, but with *Ophirina* as its only representative). Trypanosomatida are also monophyletic, while the long branch of *Perkinsela* within Prokinetoplastina, and consequently, a lack of node resolution in the lineage is expected, since this species is endosymbiotic. The monophyly of Diplonemida, Euglenida, Heterolobosea and Trypanosomatida is consistent with the consensus of previous studies (e.g. Pánek et al. (2025); Lax et al. (2021); Yazaki et al. (2017)), and supports the overall reliability of dataset construction or phylogenomics analysis workflows of our study.

Metakinetoplastina was inferred to be monophyletic. The rooting of Metakinetoplastina splits the group into two sister clades that both include neobodonids. One clade contains Trypanosomatida, Eubodonida, Parabodonida, Allobodonidae (formerly neobodonid clade '1E') and neobodonid clade '1D'; the second one contains the other two neobodonid clades that were included in the analysis: '1C', 'Nd6', and '1B' (Rhynchomonadidae). Within the first clade, Eubodonida then Parabodonida are recovered as the closest sister groups to Trypanosomatida (in agreement with

Yazaki et al. (2017)). Allobodonidae form an outgroup relative to Trypanosomatida + Eu-and Para-bodonida. The lineage of two '1D' neobodonids is sister to all other groups in the first major clade of Metakinetoplastina. In the second major clade, the single representative of 'Nd6' (KUC3-KIN1) falls sister to 1B and '1C', although the node grouping '1B' and '1C' has poor bootstrap supports in the ML analysis. The monophyly of each of '1E' (Allobodonidae), '1D', and '1C' is recovered with full support in maximum likelihood and Bayesian analyses. In our tree, the rooting of Metakinetoplastina between the four 'neobodonid' clades technically renders neobodonids paraphyletic. This observation neither verifies nor disproves any robust former inferences of the phylogenetic relationships within the 'Neobodonid' assemblage, due to the previously unreliable topology of the early branches within Metakinetoplastina (see Discussion).

Trypanosomatida constitutes a robust monophyletic lineage with *Paratrypanosoma confusum* as the sister to all other trypanosomatids, the same branching order recovered in multiple rooted phylogenies produced prior to this study Lukeš et al. (2018); Kostygov et al. (2024); Yazaki et al. (2017). The split following this basal branch divides the Trypanosomatinae (the four *Trypanosoma* species) from the rest of the subfamilies; this split has also been robustly recovered in studies over the past two decades Lukeš et al. (2018); Kostygov et al. (2024); Yazaki et al. (2017); Deschamps et al. (2011).

Within the non-Trypanosomatinae clade, *Blechomonas* is sister to all other species in our tree (consistent with Kostygov et al. (2024)). Among these species, *Blastocrithidia* is sister to Herpetomonadinae (the two *Phytomonas* species), likely due to the absence of another representative from Blastocrithidiinae in our dataset, as the branch leading to *Blastocrithidia* is long and the branch it shares with Herpetomonadinae is short. In turn, Herpetomonadinae plus *Blastocrithidia* are sister to Strigomonadinae plus Leishmaniinae.

Our topology shows Strigomonadinae (*Angomonas* and *Strigomonas*) as sister to Leishmaniinae (*Crithidia*, *Leptomonas*, *Novymonas*, *Endotrypanum*, *Porcisia* and *Leishmania*). Given our taxon sampling (in particular, the absence of *Wallacemonas* and *Sergeia*), this relationship does not contradict the one in Kostygov et al. (2024), which shows Strigomonadinae, *Sergeia* and *Wallacemonas* clade sister to Leishmaniinae. However, it is different from the topology inferred in Yazaki et al. (2017), where Herpetomonadinae (*Herpetomonas* and *Phytomonas*) form a clade with *Stringomonadinae*, and this clade is sister to Leishmaniinae.

The relationships within Leishmaniinae are identical to those in previous phylogenomic Kostygov et al. (2024) and single-gene phylogenetic Yazaki et al. (2017) analyses (though taxon sampling differs from study to study).

4 Discussion

This study represents the most extensive phylogenomic analysis of kinetoplastids thus far, especially in its taxon sampling of bodonid metakinetoplastids, and in the number of genes examined. Previous analyses specifically aiming to resolve deep kinetoplastid phylogeny have included one eubodonid, one parabodonid, and one neobodonid Deschamps et al. (2011), or one eubodonid, two parabodonids, and two neobodonids Yazaki et al. (2017). These analyses also included modest numbers of genes (<70) by contemporary standards. Low taxon sampling can be problematic in terms of phylogenetic accuracy, but can also result in skeletal phylogenetic trees that encourage misleading evolutionary inferences, even if they are topologically accurate. Our improved taxon sampling covers most of the 18S rDNA-based major subgroups of neobodonids for the first time, and arguably represents the first real test of neobodonid monophyly using a phylogenomic approach.

In addition, our analyses used a set of orthogroups (OGs) inferred *de novo* from the predicted proteomes of taxa included in our dataset (all of which belong to Discoba). There are two principal reasons why chose to create a new set of OGs, rather than analyzing existing, widely-used eukaryotic marker gene sets such as the ones proposed in Tice et al. (2021): first, OG sets which are universal for eukaryotes may lack some OGs specific to, yet shared by majority of representatives in a lineage of interest; in our case, Discoba. Second, existing lineage-specific sets of OGs, such as Euglenozoa-specific BUSCOs Manni et al. (2021), were created based on the data available for lineage representatives at the time, which, in the case of Discoba, was (and remains) heavily biased towards parasitic kinetoplastids, and as a consequence, may contain some OGs specific only to these well-studied subgroups, and lack OGs lost in these subgroups. Beyond these two main justifications, there are two other reasons that a *de novo* approach could be preferred: first, in a *de novo* approach, all input species are treated equally, in contrast to a method that adds orthologs from new species to an existing data set; in the case that either existing orthologs or new orthologs were collected with some bias, this could potentially lead to these biases affecting the topology of the final phylogenomic tree. Second, a

de novo approach should in principle be better positioned to correctly identify lineage-specific duplications that occur within the group to be analysed and separate them into different OGs.

Our new phylogenetic analyses are in fact consistent with previous studies Deschamps et al. (2011); Yazaki et al. (2017), but the much-improved taxon sampling of neobodonids, as well as the much larger number of OGs included (444 versus 64 Deschamps et al. (2011) or 43 Yazaki et al. (2017)) allowed a credible test of neobodonid monophyly, as opposed to inferring their phylogenetic placement under an implicit assumption that they represented a clade. Crucially, we find strong evidence that neobodonids are a paraphyletic assemblage, from which other metakinetoplastids descend as a monophyletic group.

Of the five subclades of neobodonids included in our analysis, Nd6, 1C, and Rhynchobodonidae (1B) form a clade, although the branching order among the three groups remains poorly resolved. The poor support is likely a result of both Nd6 and 1B being represented by a single species. The grouping of Nd6, 1C and 1B is consistent with previous phylogenomic analyses placing the 1B and 1C clades together with strong support to the exclusion of other metakinetoplastids Yazaki et al. (2017). The placement of Allobodonidae (1E) as sister to the Parabodonida, Eubodonida and Trypanosomatida clade is in harmony with the consistent and usually strongly supported bipartition between Allobodonidae and all other neobodonids in 18S rDNA trees of metakinetoplastids Packer et al. (2025); Goodwin et al. (2018); von der Heyden et al. (2004); von der Heyden and Cavalier-Smith (2005). This confirms the observation that Allobodonidae was the group most likely to be more closely related to Trypanosomatida, Eubodonida, and Parabodonida than to other *sensu lato* neobodonid clades (1A, Rhynchomonadidae, 1C, Nd6 and 1D) Goodwin et al. (2018).

The placement of 1D separate from Nd6, 1C and Rhynchomonadidae in our tree is neither supported nor contraindicated by previous analyses. There are no prior analyses of protein coding genes that include 1D and different analyses of 18S rDNA sequences show various relationships amongst 1A, Rhynchomonadidae, 1C, 1D and Nd6, with any particular set of relationships being very poorly supported Packer et al. (2025); Goodwin et al. (2018); von der Heyden et al. (2004); von der Heyden and Cavalier-Smith (2005). Very little is known at present about the biology of 1D kinetoplastids. The first isolate assigned to 1D reportedly had a spiral groove running around the cell and was informally referred to as *Cryptaulaxoides*-like von der Heyden et al. (2004); *Cryptaulaxoides* (= *Cryptaulax* Skuja 1948) refers to elongate, rapidly swimming cells similar to *Rhynchobodo* spp. (not to be confused with *Rhynchomonas*). However, subsequent isolates were listed as *Neobodo* sp. von der Heyden and Cavalier-Smith (2005) and the 1D isolates studied here (BEAP0071-"10D", BEAP0117-"13G") superficially resemble *Neobodo designis* (Figure S1), which is discussed further below. Preliminary TEM data suggests that these isolates have an eukinetoplast and a microtubular prism, like *N. designis*, however these features are found together other kinetoplastids as well (e.g. Packer et al. (2025)). There are no other published electron microscopy or genomic data.

With respect to the rest of the backbone of the kinetoplastid tree, our recovery of Prokinetoplastina and Metakinetoplastina as monophyletic sister taxa confirms previous analyses based on single genes through to several dozen genes Yazaki et al. (2017); Callahan et al. (2002); Simpson et al. (2004); Cenci et al. (2016); Tikhonenkov et al. (2021). The monophyletic group within metakinetoplastids made up of Parabodonida, Eubodonida and Trypanosomatida is also consistent with most relevant analyses of hsp90 sequences and with previous phylogenomic analyses Simpson et al. (2002, 2004); Deschamps et al. (2011); Yazaki et al. (2017). The placement of Eubodonida (i.e. *Bodo saltans*) as the closest sister to trypanosomatids has been widely assumed for decades Blom et al. (1998); Lukeš et al. (2002), but actually was not clearly resolved by 18S rDNA analyses Simpson et al. (2002); Moreira et al. (2004); Goodwin et al. (2018). Our recovery of this position confirms this relationship, however, bolstering previous phylogenomic analyses and most studies of hsp90 sequences Simpson et al. (2002, 2004); Deschamps et al. (2011); Yazaki et al. (2017).

Our sampling of significant kinetoplastid clades is not exhaustive. Within neobodonids, we have no representatives from 1A. The lack of data from 1A is notable, since this clade includes *Rhynchobodo*. Phylogenetic analyses of 1 or 2 heat shock proteins have placed a *Rhynchobodo* isolate separately from other included neobodonids, and often as the deepest branch within Metakinetoplastina, albeit statistical support was usually weak, and the taxon sampling of both neobodonids and close outgroups was very poor Simpson et al. (2004). If this position is correct, however, it implies that neobodonids are even more serially paraphyletic than our current phylogenomic results already show. A possible deep-branching position for *Rhynchobodo* is intriguing given that the recently documented free-living prokinetoplastids, *Papus* and *Apiculatomorpha* Tikhonenkov et al. (2021), are broadly similar to *Rhynchobodo* in morphology and behaviour: with a large rostrum, a spiral groove or twist to the cell, a tendency to swim through

fluid rather than along surfaces, and in being polykinetoplastidic cells with tubular extrusomes. Both of the latter features are present in the *Rhynchobodo* sp. studied by TEM only by Brugerolle Brugerolle (1985), while the clade 1A taxon *Klosteria* has tubular extrusomes, and is probably polykinetoplastidic since a eukinetoplast was absent Berney et al. (2003). It might be speculated that the common ancestor of all kinetoplastids was similar to *Papus*, *Apiculatomorpha* and *Rhynchobodo*. This speculation however, might imply multiple occasions of independent origin of *N. designis*-like morphotypes: the morphology of all representatives from 1C, Nd6 and 1D clades examined in this study is *N. designis*-like Figure S1, Figure S2. Because of this, including *Rhynchobodo* and/or other representatives of neobodonid clade 1A in future phylogenomic analyses would be especially valuable.

Here we present the first multi-gene data for a member of clade Nd6. Nd6 is one of the several clades containing isolates assigned to the morphospecies *Neobodo designis*. When 18S rDNA for isolates assigned to *N. designis* became available, it was clear that they could be highly genetically distinct from one another, and were not all closely related von der Heyden et al. (2004); von der Heyden and Cavalier-Smith (2005); Koch and Ekelund (2005). Most isolates fell within clade 1C, albeit forming several genetically distinct major subclades, while others branched elsewhere, including outside the neobodonid assemblage altogether Koch and Ekelund (2005); Goodwin et al. (2018) (see below). Nd6 was originally assigned to clade 1C, however the rest of 1C forms a moderately strongly supported clade in 18S rDNA phylogenies, while Nd6 is often not recovered as the sister to the rest of 1C at all Goodwin et al. (2018), and when it is sister, support is weak von der Heyden and Cavalier-Smith (2005). We have found Nd6 to be weakly supported as a sister to clade 1B, not branching with the rest of 1C. However, the sister position of 1B and 1C is not fully supported in ML analysis; sampling of more representatives from Nd6 and 1B clades is required.

In addition, analyses of 18S rDNA data show at least four relatively long metakinetoplastid sequences that represent multiple distinct lineages outside the recognised main groups. In unrooted trees these usually, but not always, branch between neobodonids *sensu lato*, and a parabodonid + eubodonid + trypanosomatid clan von der Heyden et al. (2004); Goodwin et al. (2018); Packer et al. (2025). These sequences could represent up to three additional clades that may be highly relevant for understanding deep-level phylogeny and evolution in metakinetoplastids, but for which we have almost no additional information at present. Most are environmental sequences, but one was from a cultivated isolate from soil that was identified as *Neobodo (Bodo) designis* Koch and Ekelund (2005), but without accompanying morphological data reported. This suggests that at least one of these clades has readily cultivable representatives that could be explored in greater detail if or when reisolated.

There is an extensive history of trying to understand aspects of the evolution of kinetoplastids across molecular phylogenies of the group Lukeš et al. (2002), and kinetoplastids were a notable early instance of using molecular phylogenetics, more or less alone, as the basis for the high-level systematics of a major group of protists Moreira et al. (2004). This study is the first time in two decades that there has been a clearly-supported update of the basic phylogenetic understanding that underpins such evolutionary inference and systematics. If our results are borne out by further analyses (some currently in progress), a revision of the high-level systematics within metakinetoplastids is required. The taxon Neobodonida appears to rationally apply to only a subset of its current composition, potentially 1C (which may be equivalent to Neobodonidae), Rhynchomonadidae, and Nd6. Our results especially highlight the importance of both Allobodonidae and clade 1D to tracing evolutionary history down the tree of kinetoplastids to trypanosomatids. Representatives of these clades would be high-priority targets for high-quality nuclear genome sequencing and annotation, explorations of mitochondrial genome organisation and extent of RNA editing, and more detailed studies of cell architecture generally. The availability of BEAP0071-"10D" and BEAP0117-"13G" isolates in culture opens these possibilities for future studies.

5 Data availability

Raw transcriptome sequencing reads for BEAP0186-"16Ckin", BEAP0230-"RhMon", BEAP0214-"G65133", BEAP0154-"D44", BEAP0117-"13G", BEAP0071-"10D", "BAB10-KIN1", "INU10-KIN1" and "KUC3-KIN1" and Gemkin are available via GenBank under the BioProject PRJNA1332963. Raw transcriptome sequencing reads for "OB23" are available via GenBank under the BioProject PRJDB40754. In addition, 18S ribosomal RNA gene sequences obtained from transcriptomic assemblies of BEAP0230-"RhMon", BEAP0214-"G65133", BEAP0154-"D44", BEAP0117-"13G", and BEAP0071-"10D", "BAB10-KIN1", "INU10-KIN1" and "KUC3-KIN1" were deposited in GenBank under IDs PZ229093-PZ229100. The 18S ribosomal RNA gene sequence from "OB23" was deposited in GenBank under IDs LC924540.

A BEAP0186-"16Ckin" cell culture (with mixed bacteria) is available in the Roscoff Culture Collection (RCC) with accession number RCC11336. The following cell cultures (with mixed bacteria) are available in the Culture Collection of Algae and Protozoa (CCAP): BEAP0230-"RhMon" (accession number CCAP 1982/5), BEAP0214-"G65133" (CCAP 1982/4), BEAP0154-"D44" (CCAP 1982/3), BEAP0117-"13G" (CCAP 1982/2), and BEAP0071-"10D" (CCAP 1982/1). All commands that were run during data analysis are stored in the file available on GitHub `https://github.com/beaplab/kineto_phylo/blob/main/MASTER_commands_V2.sh`. Assembled transcriptome contigs and predicted proteins for all newly sequenced isolates in this study are available on FigShare at `https://doi.org/10.6084/m9.figshare.30214630`.

All files generated in the process that cannot be directly obtained by running the commands from the aforementioned files (for example, the manually curated sets of orthologous genes) are available either through the GitHub repository `https://github.com/beaplab/kineto_phylo` or Figshare at `https://doi.org/10.6084/m9.figshare.30214630`.

6 Acknowledgments

We are deeply grateful to Clare Morrall for arranging access to sampling and facilities in Grenada. We thank Clara Cardelús, operating the Blanes Bay Microbial Observatory (BBMO), and Josep M. Gasol and Ramon Massana for providing access to monthly water samples from Blanes Bay. We thank Margarita Skamnelou and Àlex Gàlvez i Morante from the Biology and Ecology of Abundant Protists lab for isolating, identifying and establishing initial cultures of isolates of BEAP0154-"D44", BEAP0230-"RhMon" and BEAP0214-"G65133". We thank Dr. Keigo Yamamoto (Research Institute of Environment, Agriculture and Fisheries, Osaka Prefecture, Osaka, Japan) for arranging access to sampling and facilities in Osaka Bay. We thank Anna Karnkowska for providing environmental samples from which isolates "BAB10-KIN1", "INU10-KIN1", and "KUC3-KIN1" were isolated. We also thank Elizabeth Weston (Dalhousie University) for training JP in RNA extraction and maintaining cultures.

7 Funding

This project has received funding from the European Research Council (ERC) under the European Unions Horizon 2020 research and innovation programme (grant agreement No. 949745), from the Departament de Recerca i Universitats de la Generalitat de Catalunya (exp. 2021 SGR 00751), from grant PID2023-152955NA-I00 funded by MICIU/AEI/10.13039/501100011033 and by ERDF/EU (to D.J.R.), from the Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant 298366-2019 (to A.G.B.S.), from Grants-in-Aid for Scientific Research (No. 21H05057, No. 24K21929, and No. 25K02334) from the Japan Society for the Promotion of Science (JSPS), from the project No. 2022/47/P/NZ8/02074 (to DT) cofunded by the National Science Centre and the European Union's Horizon 2020 research and innovation programme under the Marie Skodowska-Curie grant agreement no. 945339. RK is also supported by Institute for Fermentation, Osaka (2024-G-1-011); some of the environmental samples were collected thanks to funding from the National Science Centre (Poland), grant number 2020/37/B/NZ8/01456 (to Anna Karnkowska).

Computational analyses were enabled by the supercomputing infrastructure from the Galician Supercomputing Center (CESGA; funded by the Spanish Ministry of Science and Innovation, the Galician Government and the European Regional Development Fund; ERDF).

References

Adl, S. M., Bass, D., Lane, C. E., Lukeš, J., Schoch, C. L., Smirnov, A., Agatha, S., Berney, C., Brown, M. W., Burki, F., et al. (2019). Revisions to the classification, nomenclature, and diversity of eukaryotes. *Journal of Eukaryotic Microbiology*, 66(1):4–119.

Berney, C., Fahrni, J., Petrov, N., Pawlowski, J., et al. (2003). The taxonomic position of klosteria bodomorphis gen. and sp. nov..(kinetoplastida) based on ultrastructure and ssu rrna gene sequence analysis. *Protistology*, 3(2):126–135.

Bischoff, H. W. Bold, H. C. (1963). Some soil algae from enchanted rock and related algal species. *Phycological Studies*, IV.

Blom, D., de Haan, A., van den Berg, M., Sloof, P., Benne, R., Jirku, M., and Luke, J. (1998). Rna editing in the free-living bodonid bodo saltans. *Nucleic acids research*, 26(5):1205–1213.

Brugerolle, G. (1985). Des trichocystes chez les bodonides, un caractère phylogénétique supplémentaire entre kinetoplastida et euglenida. *Protistologica (Paris. 1965)*, 21(3):339–348.

Brugerolle, G., Lom, J., Nohynkova, E., and Joyon, L. (1979). Comparaison et évolution des structures cellulaire chez plusieurs espèces de bodonidés et cryptobiidés appartenant aux genres bodo, cryptobia et trypanoplasma (kinetoplastida, mastigophora). *Protistologica*, 15:197–221.

Bushmanova, E., Antipov, D., Lapidus, A., and Prjibelski, A. D. (2019). rnaspades: a de novo transcriptome assembler and its application to rna-seq data. *GigaScience*, 8(9):giz100.

Butenko, A., Opperdoes, F. R., Flegontova, O., Horák, A., Hampl, V., Keeling, P., Gawryluk, R. M. R., Tikhonenkov, D., Flegontov, P., and Lukeš, J. (2020). Evolution of metabolic capabilities and molecular features of diplonemids, kinetoplastids, and euglenids. *BMC Biology*, 18(1):23.

Callahan, H. A., Litaker, R. W., and Noga, E. J. (2002). Molecular taxonomy of the suborder bodonina (order kinetoplastida), including the important fish parasite, ichthyobodo necator. *Journal of Eukaryotic Microbiology*, 49(2):119–128.

Capella-Gutiérrez, S., Silla-Martínez, J. M., and Gabaldón, T. (2009). trimal: a tool for automated alignment trimming in large-scale phylogenetic analyses. *Bioinformatics*, 25(15):1972–1973.

Cenci, U., Moog, D., Curtis, B. A., Tanifuji, G., Eme, L., Lukeš, J., and Archibald, J. M. (2016). Heme pathway evolution in kinetoplastid protists. *BMC Evolutionary Biology*, 16(1):109.

Chen, S., Zhou, Y., Chen, Y., and Gu, J. (2018). fastp: an ultra-fast all-in-one fastq preprocessor. *Bioinformatics*, 34(17):i884–i890.

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. *Molecular biology and evolution*, 28(1):53–58.

Emms, D. M. and Kelly, S. (2019). Orthofinder: phylogenetic orthology inference for comparative genomics. *Genome biology*, 20(1):238.

Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). Cd-hit: accelerated for clustering the next-generation sequencing data. *Bioinformatics*, 28(23):3150–3152.

Gibson, W. (2017). Kinetoplastea. In Archibald, J. M., Simpson, A. G. B., and Slamovits, C. H., editors, *Handbook of the Protists*, pages 1–50. Springer International Publishing, Cham, 2 edition.

Gigeroff, A. S., Eglit, Y., and Simpson, A. G. (2023). Characterisation and cultivation of new lineages of colponemids, a critical assemblage for inferring alveolate evolution. *Protist*, 174(2):125949.

Goodwin, J. D., Lee, T. F., Kugrens, P., and Simpson, A. G. (2018). Allobodo chlorophagus n. gen. n. sp., a kinetoplastid that infiltrates and feeds on the invasive alga codium fragile. *Protist*, 169(6):911–925.

Gruber-Vodicka, H. R., Seah, B. K. B., and Pruesse, E. (2020). phyloflash: Rapid small-subunit rrna profiling and targeted assembly from metagenomes. *mSystems*, 5(5):10.1128/msystems.00920–20.

Hide, G. (1999). History of sleeping sickness in east africa. *Clinical Microbiology Reviews*, 12(1):112–125.

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., and Vinh, L. S. (2018). Ufboot2: improving the ultrafast bootstrap approximation. *Molecular biology and evolution*, 35(2):518–522.

Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K., Von Haeseler, A., and Jermiin, L. S. (2017). Modelfinder: fast model selection for accurate phylogenetic estimates. *Nature methods*, 14(6):587–589.

Katoh, K. and Standley, D. M. (2013). Mafft multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution*, 30(4):772–780.

Koch, T. A. and Ekelund, F. (2005). Strains of the heterotrophic flagellate bodo designis from different environments vary considerably with respect to salinity preference and ssu rrna gene composition. *Protist*, 156(1):97–112.

Kostygov, A. Y., Albanaz, A. T., Butenko, A., Gerasimov, E. S., Lukeš, J., and Yurchenko, V. (2024). Phylogenetic framework to explore trait evolution in trypanosomatidae. *Trends in Parasitology*, 40(2):96–99.

Kostygov, A. Y., Karnkowska, A., Votỳpka, J., Tashyreva, D., Maciszewski, K., Yurchenko, V., and Lukeš, J. (2021). Euglenozoa: taxonomy, diversity and ecology, symbioses and viruses. *Open biology*, 11(3):200407.

Lartillot, N., Rodrigue, N., Stubbs, D., and Richer, J. (2013). Phylobayes mpi: phylogenetic reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic biology*, 62(4):611–615.

Lax, G., Kolisko, M., Eglit, Y., Lee, W. J., Yubuki, N., Karnkowska, A., Leander, B., Burger, G., Keeling, P., and Simpson, A. (2021). Multigene phylogenetics of euglenids based on single-cell transcriptomics of diverse phagotrophs. *Molecular Phylogenetics and Evolution*, 159:107088.

Li, W. and Godzik, A. (2006). Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, 22(13):1658–1659.

Lukeš, J., Butenko, A., Hashimi, H., Maslov, D. A., Votỳpka, J., and Yurchenko, V. (2018). Trypanosomatids are much more than just trypanosomes: clues from the expanded family tree. *Trends in parasitology*, 34(6):466–480.

Lukeš, J., Lys Guilbride, D., Votypka, J., Zíková, A., Benne, R., and Englund, P. T. (2002). Kinetoplast dna network: evolution of an improbable structure. *Eukaryotic cell*, 1(4):495–502.

Lukeš, J., Speijer, D., Zíková, A., Alfonzo, J. D., Hashimi, H., and Field, M. C. (2023). Trypanosomes as a magnifying glass for cell and molecular biology. *Trends in Parasitology*, 39(11):1128–1142.

Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A., and Zdobnov, E. M. (2021). Busco update: novel and streamlined workflows along with broader and deeper phylogenetic coverage for scoring of eukaryotic, prokaryotic, and viral genomes. *Molecular biology and evolution*, 38(10):4647–4654.

Minh, B. Q., Schmidt, H. A., Chernomor, O., Schrempf, D., Woodhams, M. D., Von Haeseler, A., and Lanfear, R. (2020). Iq-tree 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37(5):1530–1534.

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. *International journal of systematic and evolutionary microbiology*, 54(5):1861–1875.

Nguyen, L.-T., Schmidt, H. A., Von Haeseler, A., and Minh, B. Q. (2015). Iq-tree: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular biology and evolution*, 32(1):268–274.

Packer, J. A., Zavadska, D., Weston, E. J., Eglit, Y., Richter, D. J., and Simpson, A. G. (2025). Characterization of allobodo yubaba sp. nov. and novijibodo darinka gen. et sp. nov., cultivable free-living species of the phylogenetically enigmatic kinetoplastid taxon allobodonidae. *Journal of Eukaryotic Microbiology*, 72(1):e13072.

Pánek, T., Tice, A. K., Corre, P., Hrubá, P., Žihala, D., Kamikawa, R., Yazaki, E., Shiratori, T., Kume, K., Hashimoto, T., et al. (2025). An expanded phylogenomic analysis of heterolobosea reveals the deep relationships, non-canonical genetic codes, and cryptic flagellate stages in the group. *Molecular Phylogenetics and Evolution*, 204:108289.

Richter, D. J., Berney, C., Strassert, J. F., Poh, Y.-P., Herman, E. K., Muñoz-Gómez, S. A., Wideman, J. G., Burki, F., and de Vargas, C. (2022). Eukprot: A database of genome-scale predicted proteins across the diversity of eukaryotes. *Peer Community Journal*, 2.

Richter, D. J., Fozouni, P., Eisen, M. B., and King, N. (2018). Gene family innovation, conservation and loss on the animal stem lineage. *elife*, 7:e34226.

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., Preibisch, S., Rueden, C., Saalfeld, S., Schmid, B., Tinevez, J.-Y., White, D. J., Hartenstein, V., Eliceiri, K., Tomancak, P., and Cardona, A. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, 9(7):676–682.

Shen, W., Le, S., Li, Y., and Hu, F. (2016). Seqkit: a cross-platform and ultrafast toolkit for fasta/q file manipulation. *PloS one*, 11(10):e0163962.

Simpson, A. G., Gill, E. E., Callahan, H. A., Litaker, R. W., and Roger, A. J. (2004). Early evolution within kinetoplastids (euglenozoa), and the late emergence of trypanosomatids. *Protist*, 155(4):407–422.

Simpson, A. G., Lukeš, J., and Roger, A. J. (2002). The evolutionary history of kinetoplastids and their kinetoplasts. *Molecular Biology and Evolution*, 19(12):2071–2083.

Simpson, A. G. B. (1997). The identity and composition of the euglenozoa. *Archiv für Protistenkunde*, 148(3):318–328.

Simpson, A. G. B., Stevens, J. R., and Lukeš, J. (2006). The evolution and diversity of kinetoplastid flagellates. *Trends in Parasitology*, 22(4):168–174.

Steenwyk, J. L., Buida III, T. J., Labella, A. L., Li, Y., Shen, X.-X., and Rokas, A. (2021). Phykit: a broadly applicable unix shell toolkit for processing and analyzing phylogenomic data. *Bioinformatics*, 37(16):2325–2331.

Steverding, D. (2008). The history of african trypanosomiasis. *Parasites & Vectors*, 1(1):3.

Tashyreva, D., Simpson, A. G., Prokopchuk, G., Škodová-Sveráková, I., Butenko, A., Hammond, M., George, E. E., Flegontova, O., Záhonová, K., Faktorová, D., et al. (2022). Diplonemids–a review on" new" flagellates on the oceanic block. *Protist*, 173(2):125868.

Tice, A. K., Žihala, D., Pánek, T., Jones, R. E., Salomaki, E. D., Nenarokov, S., Burki, F., Eliáš, M., Eme, L., Roger, A. J., et al. (2021). Phylofisher: a phylogenomic package for resolving eukaryotic relationships. *PLoS Biology*, 19(8):e3001365.

Tikhonenkov, D. V., Gawryluk, R. M., Mylnikov, A. P., and Keeling, P. J. (2021). First finding of free-living representatives of prokinetoplastina and their nuclear and mitochondrial genomes. *Scientific reports*, 11(1):2946.

von der Heyden, S. and Cavalier-Smith, T. (2005). Culturing and environmental dna sequencing uncover hidden kinetoplastid biodiversity and a major marine clade within ancestrally freshwater neobodo designis. *International journal of systematic and evolutionary microbiology*, 55(6):2605–2621.

von der Heyden, S., Chao, E. E., Vickerman, K., and CAVALIER-SMITH, T. (2004). Ribosomal rna phylogeny of bodonid and diplonemid flagellates and the evolution of euglenozoa. *Journal of Eukaryotic Microbiology*, 51(4):402–416.

Xie, J., Chen, Y., Cai, G., Cai, R., Hu, Z., and Wang, H. (2023). Tree visualization by one table (tvbot): a web application for visualizing, modifying and annotating phylogenetic trees. *Nucleic acids research*, 51(W1):W587–W592.

Yazaki, E., Ishikawa, S. A., Kume, K., Kumagai, A., Kamaishi, T., Tanifuji, G., Hashimoto, T., and Inagaki, Y. (2017). Global kinetoplastea phylogeny inferred from a large-scale multigene alignment including parasitic species for better understanding transitions from a free-living to a parasitic lifestyle. *Genes & Genetic Systems*, 92(1):35–42.