Phylogenetic position and mitochondrial genome evolution of 'orphan' eukaryotic lineages

Vasily Zlatogursky[1,3,4], Vittorio Boscaro[1,3], Gordon Lax[1], Matias Wanntorp[2], Nina Pohl[2], Fabien Burki[2], Patrick J. Keeling[1]

1 Department of Botany, University of British Columbia, Vancouver, BC, Canada
2 Department of Organismal Biology, Uppsala University, Uppsala, Sweden
3 Equal contribution
4 Lead contact
* Correspondence: v.zlatogursky@gmail.com

Summary

The overall shape of the tree of eukaryotes has been brought into focus by large-scale phylogenomic analyses, highlighting a small number of highly diverse clades, or 'supergroups'[1-4]. A handful of so-called 'orphan' lineages, however, branch in uncertain positions outside known supergroups, leaving some areas of the tree unresolved. Subsets of four such orphan groups, the provorans[5], meteorids[6], hemimastigophorans[7], and telonemids[8,9], occasionally branch together in phylogenomic analyses[6,10,11], and the available mitochondrial genomes from two of them are similarly gene-rich[5,6,12], altogether raising the possibility that these lineages form a single clade with ancestrally gene-rich mitochondria. Unfortunately, they are currently underrepresented and seldom – if ever – included together in the same analysis. Despite the fact that telonemids have been studied for far longer than the others, their position remains unstable[10,13,14], and no mitochondrial genome from a telonemid culture is available. Here, we characterize a transcriptome and the mitochondrial genome of a newly described telonemid, *Microkorses curacao* gen. et sp. nov. We find telonemid mitochondria are also gene-rich, but with different genes retained than in provorans and meteorids. Our phylogenomic analyses show that provorans, hemimastigophorans, and meteorids form a well-supported but previously unrecognized supergroup, here dubbed the 'prometheids', while telonemids branch within an existing supergroup, the haptists. Overall, our analyses show that several orphan groups of eukaryotes that individually fail to be placed with confidence in the tree actually fit neatly into the emerging framework of eukaryotic relationships once more representatives are sampled.

Results and Discussion

A new telonemid

Telonemids are small, heterotrophic unicellular protists with two flagella and a mosaic of ultrastructural features that are either wholly unique or reminiscent of those from various other groups[9,15]. Environmental DNA analyses suggest that they are ubiquitous in aquatic environments[16,17], but there are currently only three formally described genera: *Telonema, Lateronema* and *Arpakorses*, with only a few species each. We here present a new genus and species of telonemids, *Microkorses curacao* gen. et sp. nov., isolated from the plankton of the Caribbean Sea around Curaçao. It was represented by nanoflagellates (5-6 µm), which were mostly sedentary under the culture conditions, but were also capable of fast spiral swimming in the water column (Fig. 1; Supplementary videos S1-3). Despite its tiny size, *M. curacao* is a voracious predator, able to efficiently clear a dense culture of bigger (7-12 µm) eukaryotic prey (the kinetoplastid, *Neobodo designis*) in a couple of weeks. *M. curacao* belongs to the TEL1 group[15] with *Telonema* and *Arpakorses* (Suppl. Fig. S1), and it is the smallest telonemid described so far; other distinguishing features include two flagella of equal length and a tendency to attach to the substratum (see Supplementary Discussion for a full taxonomic diagnosis).

Mitochondrial genome features of telonemids

Starting from a clonal culture strain (CU2384), we sequenced the complete mitochondrial genome of *M. curacao*, obtaining a circular-mapping sequence 42,908 bp long, with a G+C content of 26.5% (Supplementary Discussion, Suppl. Fig. S2, Suppl. Tables S1-S2). The mitogenome encodes 74 genes, 45 of them protein-coding; of these, 5 are unidentified open reading frames (ORFs), and 3 are ribosomal genes (*rnl*, *rns*, *rrn5*). There are genes coding for 26 tRNAs collectively capable of recognizing all 20 amino acids, but not stop codons. The telonemid mitogenome encodes 40 mitochondrial protein-coding genes out of the 74 known to theoretically exist[18,19], altogether making it considerably gene-rich: by comparison jakobids, the eukaryotic lineage with the most gene-rich mitochondrial genomes, have up to 66 protein-coding genes[20]. Richer genomes[6,18,21] can only be found in some discobans, CRuMs, cryptists, and *Malawimonas jakobiformis*, and the new supergroup described here (see below), the prometheids (Fig. 2). The mitogenome of *M. curacao* contains some rarely retained genes coding for proteins of the large ribosomal subunit, such as *rpl10*, *rpl31* and *rpl32*; this feature is shared by prometheids. However, *rpl18* is missing in prometheids but present in *M. curacao*, while *rpl20* and *rpl27* are by contrast only retained in prometheids. Both lineages share some uncommon genes, such as *tatC*, but no signs of heme biosynthesis or succinate dehydrogenase (complex II) genes were found in *M. curacao*, while being well represented in prometheids. This points to retentions of different sets of mitogenome-encoded proteins in prometheids and telonemids. The *M. curacao* mitogenome content is relatively similar to that of centrohelids, especially the ribosomal protein set (Fig. 2), but it is considerably more gene-rich than the mitogenome of the other major group of Haptista, the haptophytes (again, especially in ribosomal protein genes). Overall, the extended repertoire of mitochondrial genes in telonemids does not suggest any especially close relationship to meteorids or provorans (whose mitogenome do instead resemble each other's[6]), indicating that the rich mitochondrial genomes are not an ancestrally shared character of these 'orphan' groups.

Placement of telonemids and other 'orphan' groups in the tree of eukaryotes

Telonemids have been poorly represented in phylogenomic datasets until recently[14], while other 'orphan' lineages, specifically provorans, meteorids, and hemimastigophorans, were molecularly characterized in rapid succession[5-7] and hence never all included in the same phylogenomic tree. To analyse these orphan groups together in the context of eukaryote evolution, we built two alignment matrices using 278 protein-coding genes: one including 89 taxa, chosen to optimize both completeness and taxonomic breadth, and a much larger one with 433 taxa, vastly increasing the number of representatives of all eukaryotic groups. The 89-taxa Maximum-Likelihood (ML) tree was inferred using two complex models (LG+C60+G4 and LG+UDM64+G4) which produced identical topologies (Fig. 3); the 433-taxa ML tree was inferred with a less parameter-rich model (LG+C20+G4) and produced a similar result (Suppl. Fig. S3). In particular, all analyses showed telonemids branching with Haptista, although the exact relationship between telonemids and the two major haptist subgroups, haptophyte algae and the heterotrophic centrohelids, could not be determined with confidence. Provorans, meteorids, and hemimastigophorans always formed a single clade, here dubbed 'prometheids' from the initial letters of their names and after the mythological Greek titan, as a nod to the ancient origin of their most recent common ancestor. Prometheids and haptists in turn branched within a very diverse and well-recognized clade also containing SAR, Archaeplastida, and Cryptista (the latter two always being sister groups in all our analyses).

We then assessed the stability of the two nodes of interest: the monophyly of telonemids and haptists, and that of prometheids. To determine the impact of substitution rate heterogeneity across sites, we repeated the 89-taxa dataset ML analyses on matrices with incremental removal of fast-evolving sites; the monophyly of prometheids was always recovered with very high support; support for the monophyly of telonemids and haptists increased when the most extreme fast-evolving sites were removed, but eroded or disappeared once 20% or more of the information was lost (Fig. 3). The two nodes were also recovered, with similar or higher support values, in matrices that only included the taxa with the highest data completeness (Suppl. Fig. S4), and when removing faster-evolving taxa (Suppl. Fig. S5). Telonemids were previously found to branch in several different parts of the tree of eukaryotes, including recently as sister to SAR in the 'TSAR' hypothesis[14]. To test this and other proposed or plausible placements, we built ML trees where the telonemids were constrained to cluster with either SAR, prometheids, or in a clade with Archaeplastida and Cryptista. The 'TSAR' scenario was not rejected by Approximately Unbiased tests (*p*-values: 0.25 and 0.07 with the LG+C60+G4 and LG+UDM64+G4 models, respectively), while the clustering of telonemids with prometheids was rejected by the C60 model (*p*-value: 0.06 with the LG+UDM64+G4 model); a relationship with Archaeplastida and Cryptista was rejected by both models (Fig. 3).

The monophyly of prometheids is not a complete surprise. While no tree with enough representatives was reconstructed before, meteorids and hemimastigophorans have been shown to branch together[6] despite their completely different morphologies and modes of locomotion. Meteorids are non-flagellated, gliding protists with two characteristic lateral 'arms' swinging back and forth, while hemimastigophorans bear multiple flagella (or 'cilia') arranged in two lateral rows (kineties) – and yet, while sister groups, they are approximately as divergent as the two superficially similar lineages of provorans (nibblerids and nebulids) are from each other (Fig. 3 and Suppl. Fig. S3). Provorans are biflagellated protists that voraciously feed on other

eukaryotes, using poisonous extruding organelles (which have been suggested to have homologs in hemimastigophorans) and small but strong sub-cellular 'jaws'[22]. While more data is needed, especially from the mitochondria of hemimastigophorans, it is reasonable to assume that the prometheid ancestor had a gene-rich mitogenome, as is still the case in its extant descendants.

The placement of telonemids is much more controversial. Interestingly, the first phylogenomic analysis of telonemids had them branch with centrohelids[13], but this association was not observed in later trees, which recovered various topologies of which 'TSAR' is only one[10,11]. Multiple authors have remarked that different morphological features of telonemids could point to different relatives[14,15] or to none at all if they evolved independently, as is the case for their unique cytoskeleton[23]. Telonemid sequences are not especially fast-evolving, so the reason behind the erratic behaviour of telonemids in phylogenomic tree must lie elsewhere, for example in their sparse sampling and the lack of close relatives.

Conclusions

While the catalogue of eukaryotic diversity keeps growing, we show here that several orphan groups, until now in uncertain relationships with each other and with known supergroups, can be placed with confidence in the tree when analysed together. A few years ago, three other orphan groups that lacked unifying morphological features were recognized based on molecular phylogenies to form a monophyletic group, CRuMs[24]. The prometheids are even more varied; for example, each of them represents one of the three main forms of locomotion among protists: ciliary swimming, flagellar swimming, and gliding. Yet our analyses show that they form a strongly supported single clade. Their common ancestor existed so long ago that it is not surprising its descendants have little in common (not to mention, more prometheid diversity will undoubtedly be uncovered in the future) - and yet, they all share a similarly gene-rich mitochondrial genome. We also show that another elusive group, the telonemids, also has a gene-rich mitogenome but is not related to the prometheids, according to both comparative genomics and phylogenomics. The telonemids' most likely placement is within the well-established supergroup, Haptista, based on the inclusion of many representatives for each group of eukaryotes, multiple models, and multiple taxon selections, including one particularly extensive phylogenomic tree (Suppl. Fig. S3). This conclusion is nevertheless still not definitive, since one proposed alternative, the so-called 'TSAR' hypothesis, could not be rejected. Removing these four lineages from the category of 'orphans' further solidifies a single and comprehensive tree that will be a useful tool as we interpret the incredible diversity of eukaryotic cells.

Resource availability

Lead contact

Requests for further information and resources can be directed to the lead contact, Vasily Zlatogursky (v.zlatogursky@gmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

- The *Microkorses curacao* mitochondrial genome sequence and transcriptomic raw reads have been deposited on GenBank as [accession number pending] and are publicly available as of the date of publication.
- This paper does not report original code.
- Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

Acknowledgements

Authors are thankful to Samuel Livingston for assistance with electron microscopy, Mahara Mtawali for collecting the sample, and Adarsh Hurdeal for helping with imaging. This work was supported by grants from the Natural Sciences and Engineering Research Council of Canada (NSERC 2019-03994) and the Gordon and Betty Moore Foundation (https://doi.org/10.37807/GBMF9201) to PJK, and by grants from the European Research Council (ERC consolidator grant 101044505), the Swedish Research Council VR (2021-04055), and Science for Life Laboratory to FB.

Author contributions

V.Z., V.B., and P.J.K. designed the project. V.Z. isolated and cultured the organisms, performed microscopy and molecular biology experiments, and analyzed mitochondrial data. V.B. performed phylogenetic analyses, with assistance from G.L. M.W., N.P., and F.B. curated the starting PhyloFisher database. V.Z., V.B., and P.J.K. wrote the original draft of the paper. All authors contributed to the final manuscript.

Declaration of interests

The authors declare no competing interests.

STAR methods

Detailed methods are provided in the online version of this paper and include the following:

Isolation, culturing, and morphology of *Microkorses curacao* gen. n., sp. n.
Microkorses curacao strain CU2384 was isolated from a plankton-tow collected water sample taken from the Caribbean Sea at the water factory dive site in Willemstad, Curaçao on October 10th, 2023 (latitude/longitude coordinates: 12.108993807580502, -68.95353029793979). The salinity measured in the sample was 36‰. The sample was transported to the laboratory and inoculated in a 90 mm plastic Petri dish. The dish was incubated at room temperature for two weeks with addition of 0.025% wheat grass extract to promote bacterial growth. The incubated

dish was examined using a Zeiss AxioVert A1 inverted microscope, equipped with phase contrast (PC). The living cells of *Microkorses* were manually picked using tapered glass Pasteur pipettes and inoculated individually into wells of a plastic 96-well plate, containing 200 µl of actively growing *Neobodo designis* culture in 0.025% wheat grass extract in 30 ppt artificial sea water. One well, where a clone of cells was noted to be successfully growing, was fully transferred to a Petri dish with the same *N. designis* culture and the obtained clonal culture was further maintained by bi-weekly reinoculation of 1 ml of old culture onto a fresh dish with the prey.

Light microscopy was performed in plastic Petri dishes using a Zeiss AxioVert A1 with PC or on temporary object slide preparations on the same microscope, but also on a Zeiss Axioplan 4 upright microscope, equipped with differential interference contrast (DIC) optics. The photographs and videos were captured with a Sony alpha-7R Mark 3 camera. For transmission electron microscopy the cells were fixed using high-pressure freezing (HPF). First, the cells were concentrated by centrifugation at 560 *g* for 10 minutes, the supernatant was then discarded and 1.4 µl of the concentrated cells was applied to brass planchets and immediately frozen using a Leica HPM100 HPF system. The frozen planchets were transferred to cryovials containing 2% OsO_4 and 0.1% uranyl acetate (UA) in anhydrous acetone and processed in an automatic freeze-substitution system (Leica AFS2) according to the following schedule: 96 hours at -90 °C, heated to -50 °C over 12 hours, left at -50 °C for 8 hours, heated to -20 °C over 12 hours, left at -20 °C for 8 hours, then heated to 22 °C for 12 hours. Samples were rinsed 3 times in anhydrous acetone, and then infiltrated in an ascending graded series of Spurr's resin in acetone (20%, 40%, 60%, 80%, then three infiltrations of 100%) for a minimum of 3 hours at each step. Samples were polymerized at 60 °C for 48 hours. Sections were cut from Spurr's embedded blocks using a Leica UC7 Ultramicrotome and collected on formvar-coated copper grids, post-stained with 2% aqueous UA and lead citrate and imaged using a Tecnai Spirit TEM operating at 80 kV with a DVC1500M side-mounted camera. The size of the microscopic structures was measured on micrographs using ImageJ[25] version 1.54f.

Transcriptome and mitogenome sequencing and assembly.

In order to obtain single-cell transcriptomes, the cells were manually picked and frozen in 0.2 mL plastic tubes containing 2 µl of the lysis buffer. Five cells were combined in each tube in order to increase the cDNA yield. The downstream processing of the cells was performed following the Smart-seq2 protocol[26,27] using 24 PCR cycles. For genomic sequencing, the cells were concentrated at 2,700 *g* for 10 minutes, then the total genomic DNA was extracted using the DNeasy Plant Pro kit (Qiagen). Both cDNA and genomic DNA were then sequenced with Illumina. Illumina (Nextera XT) libraries were prepared and sequenced on the Illumina NovaSeq platform at the Sequencing and Bioinformatics Consortium at the University of British Columbia (raw reads are available in the Sequence Read Archive under project to be included prior to publication). The raw reads were error-corrected using Rcorrector[28] version 1.0.4 and trimmed using Trimmomatic[29] version 0.39 with the options LEADING:5 TRAILING:5 SLIDINGWINDOW:5:16 MINLEN:60, specifying Nextera adapters, Illumina-specific primers and smartseq2 primers for removal. Transcriptomes were assembled with SPAdes[30] version 3.15.1 using reverse and unpaired Illumina reads (transcriptomes were assembled with the --rna flag, genomes with the --meta flag). Protein-coding sequences were predicted from the assemblies using Transdecoder[31] version 5.5.0.

To reconstruct the *M. curacao* mitochondrial genome, the genomic assembly was blasted using barrnap-obtained mitochondrial SSU sequence as a query, which resulted in the recovery of a single circular contig. Annotation was performed using the MFannot webserver[32], and the obtained .sqn file was converted to GenBank format using asn2gb (https://ftp.ncbi.nlm.nih.gov/asn1-converters/by_program/asn2gb/). In parallel, the same fasta file was annotated with PROKKA[33] version 1.14.6 using the --kingdom Mitochondria flag. The MFannot file was manually corrected based on the PROKKA output wherever it was better performing. The genome map was generated using OGDRAW[34] version 1.3.1.

Phylogenetic analyses.

SSU rRNA gene sequences were extracted from the obtained transcriptome using barrnap version 0.9 (https://github.com/tseemann/barrnap). Other telonemid sequences were downloaded from the PR2 database[35] and manually supplemented with some additional sequences individually downloaded from GenBank. The resulting fasta file was clustered at 99% to avoid redundancy using usearch[36] version 12, aligned using mafft[37] version 7.48 and trimmed using trimAl[38] version 1.4.r22 with a gap threshold of 0.3 and similarity threshold of 0.001. The obtained file was used for Maximum-Likelihood (ML) analysis using RAxML-NG[39] version 1.1.0 under the GTRCAT model and 1,000 non-parametric bootstrap replicates. Visualisation of this and subsequent trees was done with iTOL[40] version 6.9.1, FigTree version 1.4.4 (https://tree.bio.ed.ac.uk/software/figtree/), InkScape version 1.2, and Adobe Illustrator version 27.5.

The newly obtained telonemid transcriptome and other genomic and transcriptomic data recently made available (mostly collected from EukProt[41]) were added to a previous dataset[4], formatted for PhyloFisher[10]. 320 single-gene trees of 1,129 taxa were inferred with the PhyloFisher sgt_constructor.py pipeline, then curated manually to remove paralogs and suspected contaminants. At the end of this preliminary step, genes present in less than 50% of the taxa, as well as a few genes whose tree topology was too ambiguous or too complex (for example, due to large duplication events) to confidently curate, were removed, leaving 278 genes. After removing bacterial and archaeal data, taxa with less than 10% site coverage, the notoriously fast-evolving Ascetosporea[42], and nucleomorph data, 793 taxa remained. Preliminary phylogenomic trees were built on this comprehensive dataset. 17 very long-branching taxa were further removed; many others were pruned due to redundancy (multiple taxa from the same genus or species, and/or closely related to each other; in each case, the most complete representative was chosen). This left the 433 taxa used in the large dataset analyzed in the paper; 89 of these, chosen to maintain a wide taxonomic breadth while maximizing completeness (measured as both gene- and site-percentage representation in the supermatrix) whenever possible, constituted the smaller dataset. The alignment matrix in both datasets included 278 genes and 73,081 sites.

All multi-gene ML inferences were performed on IQ-TREE[43] version 2.2.0. The 433-taxa dataset was used to infer a ML tree with the LG+C20+G4 model, running 1,000 ultrafast bootstraps for node support. ML trees were generated for the 89-taxa dataset using the more complex LG+C60+G4 and LG+UDM64+G4 models, running 1,000 ultrafast bootstraps for node support. To assess the impact of sites with higher substitution rates, the 5% fastest-evolving sites were removed from the original 89-taxa alignment using the fast_site_removal.py script, in incremental steps up to 30%. Ultrafast bootstrapped ML trees were run on each submatrix, extracting the boostrap values of nodes of interest. Additional ML inferences based on both the

LG+C60+G4 and the LG+UDM64+G4 models were run on datasets including only taxa with 75% or more gene coverage (81 out of the original 89) and only taxa with shorter branches (78 out of 89). In unrooted trees (as all ours are, given the lack of consensus on the position of the root of eukaryotes) there is no single way to define the total branch length of a taxon, so the choice of which taxa to remove was based on the unevenness branch lengths between related taxa; this included all metamonads, *Neobodo designis*, all retarians (i.e. *Astrolonche serrata*, *Reticulomyxa filosa*, and *Elphidium margaritaceum*), *Blastocystis hominis*, *Tetrahymena termophila*, and *Cryptosporidium muris*.

Constrained trees were built to test alternative topologies on the 89-taxa dataset. In each case, a few nodes were fixed before running a full phylogenetic analysis with the LG+C60+G4 model to obtain the best topology. Specifically, the three constrained trees were run with the following nodes set: (i) telonemids (monophyletic) as sister group of SAR (monophyletic); (ii) telonemids (monophyletic) as sister group of prometheids (monophyletic); (iii) telonemids (monophyletic), Archaeplastida (monophyletic), and Cryptista (monophyletic) forming a single clade (but in no pre-defined relationship with each other). Notably, the constrained position of telonemids was the only difference between constrained trees and the unconstrained ML tree. Approximately Unbiased (AU) tests were then run in IQ-TREE (using 10,000 RELL replicates) to determine which of these three alternative topologies could be significantly rejected, according to both the LG+C60+G4 model and the LG+UDM64+G4 model.

Supplemental information

Suppl. Discussion. Taxonomic diagnosis and further consideration on mitochondrial gene features.

Suppl. Fig. S1. Maximum-Likelihood tree based on SSU rRNA gene sequences, inferred using 1,736 nucleotide positions from 120 telonemids and 10 diverse eukaryotes as an outgroup, under the GTRCAT model, with 1,000 bootstrap replicates. The new *Microkorses curacao* species is shown in red. The SSU sequences corresponding to environmental mitogenomes are shown in blue. Morphologically characterized taxa are shown in green.

Suppl. Fig. S2. The map of the *Microkorses curacao* mitochondrial genome with genes colour coded as to their function. Genes on the outside of the circle are transcribed in the clockwise direction.

Suppl. Fig. S3. Maximum-Likelihood phylogenetic tree of 433 representative eukaryotes. The character matrix included 278 protein-coding genes and 73,081 sites, and the tree was inferred under the LG+C20+G4 model, running 1,000 ultrafast bootstraps for node support; 100% bootstrap scores from both models are shown as black dots. The position of the root was chosen arbitrarily. All the main features of the 89-taxa tree are confirmed in this larger tree, including the branching of telonemids with Haptista, the monophyly of prometheids, diaphoretickes, and Archaeplastida+Cryptista.

Suppl. Fig. S4. Maximum-Likelihood phylogenetic tree of a subset of 81 taxa from Fig. 1, selected based on their high (75% or more) gene coverage. Tree and the first ultrafast bootstrap value for each node were inferred with the LG+C60+G4 model. The second bootstrap value for each node was calculated with the LG+UDM64+G4 model. Bootstrap values below 85% are

omitted, and 100% bootstrap scores from both models are shown as black dots. The completeness of each taxon's genome or transcriptome is shown on the left, as a percentage of represented genes (top bars) and non-gap sites (bottom bars). The position of the root was chosen arbitrarily. The topology corresponds to the topology in the main analyses.
Suppl. Fig. S5. Maximum-Likelihood phylogenetic tree of a subset of 78 taxa from Fig. 1, after removing 11 taxa with longer branches. Tree and the first ultrafast bootstrap value for each node were inferred with the LG+C60+G4 model. The second bootstrap value for each node was calculated with the LG+UDM64+G4 model. Bootstrap values below 85% are omitted, and 100% bootstrap scores from both models are shown as black dots. The completeness of each taxon's genome or transcriptome is shown on the left, as a percentage of represented genes (top bars) and non-gap sites (bottom bars). The position of the root was chosen arbitrarily. The topology corresponds to the topology in the main analyses.
Suppl. Fig. S6. The comparison of the gene content and synteny in *Microkorses curacao* and three previously obtained SAG-based mitogenomes. Genome maps are shown as linear for clarity. The vertical lines link corresponding genes.
Suppl. Video S1. *Microkorses curacao* gen. et sp. nov. imaged using inverted Zeiss Axiovert A1 microscope with phase contrast on temporary object slide with coverslip (same slide as in Suppl. video S3).
Suppl. Video S2. *Microkorses curacao* gen. et sp. nov. imaged using inverted Zeiss Axiovert A1 microscope with phase contrast in a Petri dish.
Suppl. Video S3. *Microkorses curacao* gen. et sp. nov. imaged using an upright Zeiss Axioplan 4 microscope with differential interference contrast on a temporary object slide with coverslip (same slide as in Suppl Video S1).
Suppl. Table S1. Comparison of the tRNA gene content in *M. curacao* and three previously obtained SAG-based mitogenomes.
Suppl. Table S2. A table supplementing Fig. 2 and showing mitogenome completeness across eukaryotes.

Figure legends

Figure 1. Cell morphology of *Microkorses curacao* gen. et sp. nov.
A. General view of the cell. Light microscopy, differential interference contrast. The scale bar stands for 5 µm. B. The mitochondrion with tubular cristae is visible in this transmission electron microscopy cross-section. The scale bar stands for 500 nm.

Figure 2. Eukaryotic mitogenomes coding capacity
Comparison of presence (circle) / absence (no circle) of protein-coding genes in mitochondrial genomes of eukaryotes (updated and corrected when necessary from [6,18,21]). The *Microkorses* branch is shown in red, prometheids are in bold and in black. Abbreviations: CI-CV - Respiratory complexes I-V; CytC - Cytochrome C.

Figure 3. The position of telonemids and 'prometheids' in the tree of eukaryotes
Multigene Maximum-Likelihood phylogenetic tree of eukaryotes, showing that telonemids belong to Haptista while hemimastigophorans, provorans, and meteorids form a single clade, the 'prometheids'. The character matrix for 89 representative eukaryotes was built with 278 protein-

coding genes and 73,081 sites. The tree and the first ultrafast bootstrap value for each node were inferred with the LG+C60+G4 model. The second bootstrap value for each node was calculated with the LG+UDM64+G4 model. Bootstrap values below 85% are omitted, and 100% bootstrap scores from both models are shown as black dots. For space reasons, the four terminal branches in Metamonada and one terminal branch in Discoba were shortened by 50%. The completeness of each taxon's genome or transcriptome is shown on the left, as a percentage of represented genes (top bars) and non-gap sites (bottom bars). Alternative positions for telonemids assessed with AU tests are shown (asterisks in the two boxes correspond to a rejection with the LG+C60+G4 model, on the left, and the LG+UDM64+G4, on the right). Graphs next to the Haptista and prometheid clades show bootstrap value changes when fast-evolving sites are removed (the absence of a datapoint denotes that the node was not recovered with the corresponding matrix). Note that the position of the root was chosen arbitrarily.

References

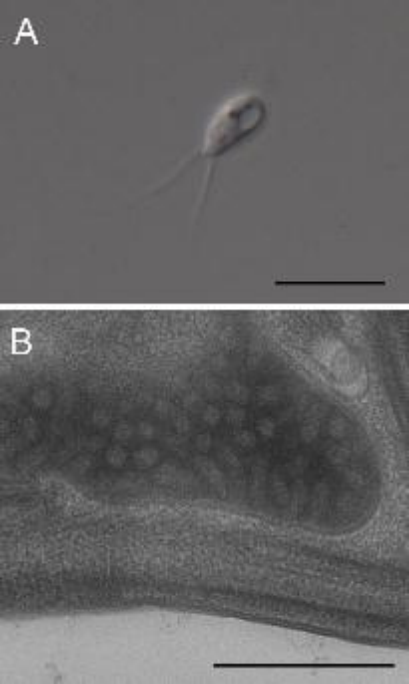
1. Brown, M.W., Heiss, A.A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A. K., Shiratori, T., Ishida, K.-I., Hashimoto, T., Simpson, A.G.B., and Roger, A.J. (2018). Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. Genome Biol. Evol. *10*, 427-433. https://doi.org/10.1093/gbe/evy014.
2. Keeling, P.J., and Burki, F. (2019). Progress towards the Tree of Eukaryotes. Curr. Biol. *29*, R808–R817. https://doi.org/10.1016/j.cub.2019.07.031.
3. Strassert, J.F.H., Irisarri, I., Williams, T.A., and Burki, F. (2021). A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. Nat. Comm. *12*, 1879. https://doi.org/10.1038/s41467-021-22044-z.
4. Schön, M.E., Zlatogursky, V.V., Singh, R.P., Poirier, C., Wilken, S., Mathur, V., Strassert, J.F.H., Pinhassi, J., Worden, A.Z., Keeling, P.J., Ettema, T.J.G., Wideman J.G., and Burki, F. (2021). Single cell genomics reveals plastid-lacking Picozoa are close relatives of red algae. Nat. Comm. *12*, 6651. https://doi.org/10.1038/s41467-021-26918-0.
5. Tikhonenkov, D.V., Mikhailov, K.V., Gawryluk, R.M.R., Belyaev, A.O., Mathur, V., Karpov, S.A., Zagumyonnyi, D.G., Borodina, A.S., Prokina, K.I., Mylnikov, A.P., Aleoshin, V.V., and Keeling, P.J. (2022). Microbial predators form a new supergroup of eukaryotes. Nature *612*, 714-719. https://doi.org/10.1038/s41586-022-05511-5.
6. Eglit, Y., Shiratori, T., Jerlström-Hultqvist, J., Williamson, K., Roger, A.J., Ishida, K.-I., and Simpson, A.G.B. (2024). *Meteora sporadica*, a protist with incredible cell architecture, is related to Hemimastigophora. Curr. Biol. *34*, 451-459. https://doi.org/10.1016/j.cub.2023.12.032.
7. Lax, G., Eglit, Y., Eme, L., Bertrand, E.M., Roger, A.J., and Simpson, A.G.B. (2018). Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. Nature *564*, 410-414. https://doi.org/10.1038/s41586-018-0708-8.
8. Griessmann, K. (1913). Über marine Flagellaten. Arch. Protistenk. 32, 1-78.
9. Shalchian-Tabrizi, K., Eikrem, W., Klaveness, D., Vaulot, D., Minge, M.A., Le Gall, F., Romari, K., Throndsen, J., Botnen, A., Massana, R., Thomsen, H.A., and Jakobsen, K.S. (2006). Telonemia, a new protist phylum with affinity to chromist lineages. Proc. Biol. Sci. *273*, 1833-1842. https://doi.org/10.1098/rspb.2006.3515.
10. Tice, A.K., Žihala, D., Pánek, T., Jones, R.E., Salomaki, E.D., Nenarokov, S., Burki, F., Eliáš, M., Eme, L., Roger, A.J., Rokas, A., Shen, X.-X., Strassert, J.F.H., Kolísko, M., and Brown, M.W. (2021). PhyloFisher: A phylogenomic package for resolving eukaryotic relationships. PLoS Biol. *19*, e3001365. https://doi.org/10.1371/journal.pbio.3001365.
11. Williamson, K., Eme, L., Baños, H., McCarthy, C.G.P., Susko, E., Kamikawa, R., Orr, R.J.S., Muñoz-Gòmez, S.A., Minh, B.Q., Simpson, A.G.B., and Roger, A.J. (In press). A robustly rooted tree of eukaryotes reveals their excavate ancestry. Nature.
12. Janouškovec, J., Tikhonenkov, D.V., Burki, F., Howe, A.T., Rohwer, F.L., Mylnikov, A.P., and Keeling, P.J. (2017) A new lineage of eukaryotes illuminates early mitochondrial genome reduction. Curr. Biol. *27*, 3717–3724. https://doi.org/10.1016/j.cub.2017.10.051.
13. Burki, F., Inagaki, I., Bråte, J., Archibald, J.M., Keeling, P.J., Cavalier-Smith, T., Sakaguchi, M., Hashimoto, T., Horak, A., Kumar, S., Klaveness, D., Jokobsen, K.S.,

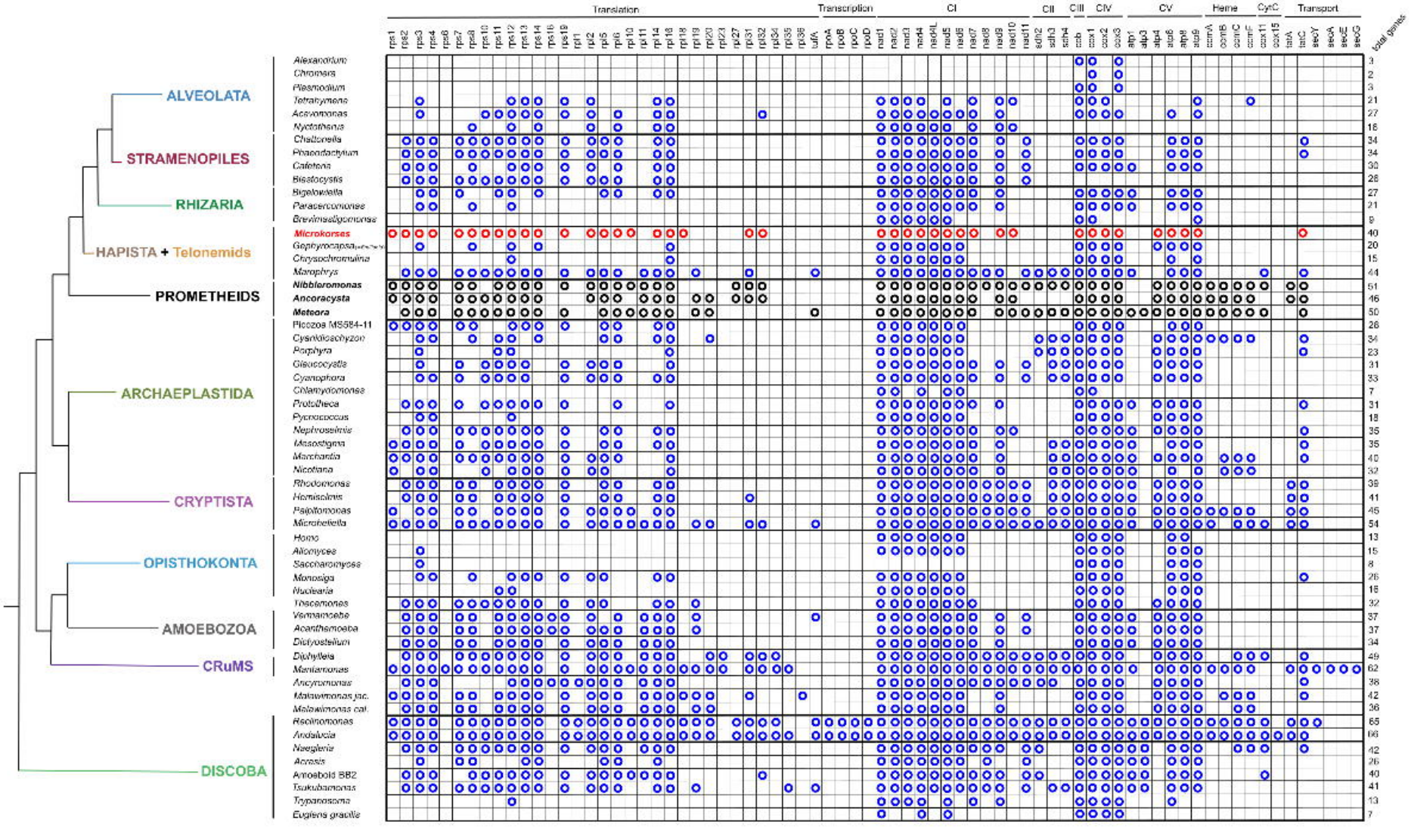
Pawlowski, J., and Shalchian-Tabrizi, K. (2009). Large-scale phylogenomic analyses reveal that two enigmatic protist lineages, Telonemia and Centroheliozoa, are related to photosynthetic chromalveolates. Genome Biol. Evol. *1*, 231-238. 10.1093/gbe/evp022.

14. Strassert, J.F.H., Jamy, M., Mylnikov, A.P., Tikhonenkov, D.V., and Burki, F. (2019). New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. Mol. Biol. Evol. *36*, 757-765. https://doi.org/10.1093/molbev/msz012.
15. Tikhonenkov, D.V., Jamy, M., Borodina, A.S., Belyaev, A.O., Zagumyonnyi, D.G., Prokina, K.I., Mylnikov, A.P., Burki, F., and Karpov, S.A. (2022). On the origin of TSAR: morphology, diversity and phylogeny of Telonemia. Open Biol. *12*, 210325. https://doi.org/10.1098/rsob.210325.
16. Bråte, J., Klaveness, D., Rygh, T., Jakobsen, K.S., and Shalchian-Tabrizi, K. (2010). Telonemia-specific environmental 18S rDNA PCR reveals unknown diversity and multiple marine-freshwater colonizations. BMC Microbiol. *10*, 168. https://doi.org/10.1186/1471-2180-10-168.
17. Boukheloua, R., Mukherjee, I., Park, H., Šimek, K.,Kasalický, V., Ngochera, M., Hans-Peter Grossart, H.-P., Picazo-Mozo, A., Camacho, A., Cabello-Yeves, P.J., et al. (2024). Global freshwater distribution of *Telonemia* protists. ISME J. *18*, wrae177. https://doi.org/10.1093/ismejo/wrae177.
18. Moreira, D., Blaz, J., Kim, E., and Eme, L. (2024). A gene-rich mitochondrion with a unique ancestral protein transport system. Curr. Biol. *34*, 3812-3819. https://doi.org/10.1016/j.cub.2024.07.017.
19. Wideman, J.G., Monier, A., Rodríguez-Martínez, R., Leonard, G., Cook, E., Poirier, C., Maguire, F., Milner, D.S., Irwin, N.A.T., Moore, K., et al. (2019). Unexpected mitochondrial genome diversity revealed by targeted single-cell genomics of heterotrophic flagellated protists. Nat. Microbiol. *5*, 154-165. https://doi.org/10.1038/s41564-019-0605-4.
20. Gray, M.W., Burger, G., Derelle, R., Klimeš, V., Leger, M.M., Sarrasin, M., Vlček, Č, Roger, A.J., Eliáš, M., and Lang, B.F. (2020). The draft nuclear genome sequence and predicted mitochondrial proteome of *Andalucia godoyi*, a protist with the most gene-rich and bacteria-like mitochondrial genome. BMC Biol. *18*, 22. https://doi.org/10.1186/s12915-020-0741-6.
21. Kamikawa, R., Shiratori, T., Ishida, K.-I., Miyashita, H., and Roger, A.J. (2016). Group II intron-mediated *trans*-splicing in the gene-rich mitochondrial genome of an enigmatic eukaryote, *Diphylleia rotans*. Genome Biol. Evol. *8*, 458-466. https://doi.org/10.1093/gbe/evw011.
22. Belyaev, A.O., Karpov, S.A., Keeling, P.J., and Tikhonenkov, D.V. (2024). The nature of ‘jaws’: a new predatory representative of Provora and the ultrastructure of nibbling protists. Open Biol. *14*, 12. https://doi.org/10.1098/rsob.240158.
23. Yabuki, A., Eikrem, W., Takishita, K., and Patterson, D.J. (2013). Fine structure of *Telonema subtilis* Griessmann, 1913: a flagellate with a unique cytoskeletal structure among eukaryotes. Protist *164*, 556-569. https://doi.org/10.1016/j.protis.2013.04.004.
24. Brown, M.W., Heiss, A.A., Kamikawa, R., Inagaki, Y., Yabuki, A., Tice, A.K., Shiratori, T., Ishida, K.-H., Hashimoto, T., Simpson, A.G.B., and Roger, A.J. (2018). Phylogenomics places orphan protistan lineages in a novel eukaryotic super-group. Genome Biol. Evol. *10*, 427-433. https://doi.org/10.1093/gbe/evy014

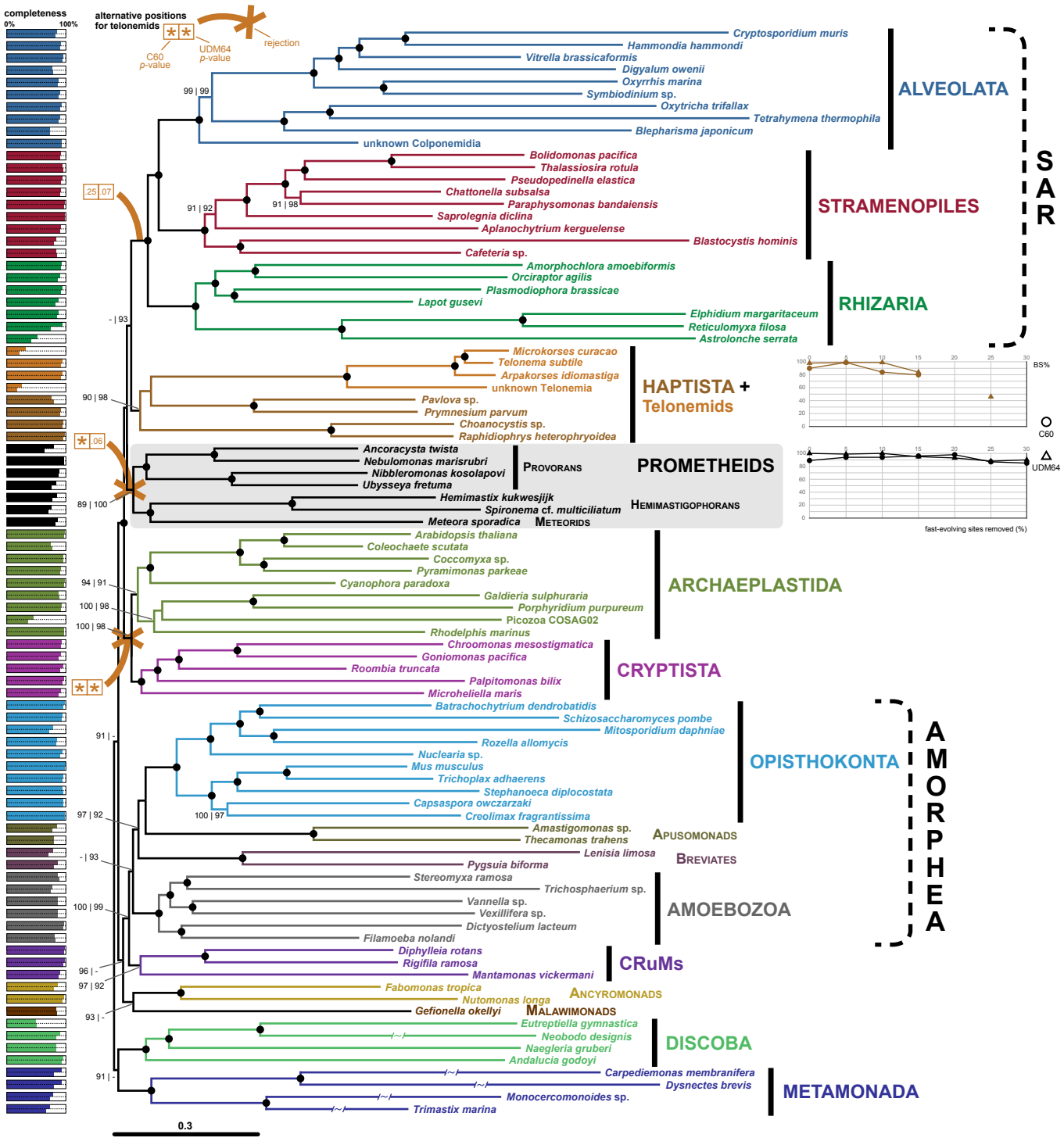
25. Schneider, C.A., Rasband, W.S., and Eliceiri, K.W. (2012). NIH Image to ImageJ: 25 years of image analysis. Nat. Methods 9, 671-675. https://doi.org/10.1038/nmeth.2089.
26. Picelli, S., Faridani, O.R., Björklund, A.K., Winberg, G., Sagasser, S., and Sandberg, R. (2014). Full-length RNA-seq from single cells using Smart-seq2. Nat. Protoc. 9, 171-181. https://doi.org/10.1038/nprot.2014.006.
27. Kolisko, M., Boscaro, V., Burki, F., Lynn, D.H., and Keeling, P.J. (2014) Single-cell transcriptomics for microbial eukaryotes. Curr. Biol. *24*, R1081-1082. https://doi.org/10.1016/j.cub.2014.10.026.
28. Song, L., and Florea, L. (2015). Rcorrector: efficient and accurate error correction for Illumina RNA-seq reads. GigaScience *4*, 48. https://doi.org/10.1186/s13742-015-0089-y.
29. Bolger, A.M., Lohse, M., and Usadel, B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30: 2114-2120. https://doi.org/B10.1093/bioinformatics/btu170.
30. 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, giz100. https://doi.org/10.1093/gigascience/giz100.
31. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J., Couger, M.B., Eccles, D., et al. (2013). De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. *8*, 1494-1512. https://doi.org/10.1038/nprot.2013.084.
32. Lang, B.F., Beck, N., Prince, S., Sarrasin, M., Rioux, P., and Burger, G. (2023). Mitochondrial genome annotation with MFannot: a critical analysis of gene identification and gene model prediction. Front. Plant Sci. *14*, 1222186. https://doi.org/10.3389/fpls.2023.1222186.
33. Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. Bioinformatics *15*, 2068-2069. https://doi.org/10.1093/bioinformatics/btu153.
34. Greiner, S., Lehwark, P., and Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. Nucleic Acids Res. *46*, W59-W64. https://doi.org/10.1093/nar/gkz238.
35. Guillou, L., Bachar, D., Audic, S., Bass, D., Berney, C., Bittner, L., Boutte, C., Burgaud, G., de Vargas, C., Decelle, J., et al. (2013). The Protist Ribosomal Reference database (PR^2): a catalog of unicellular eukaryote Small Sub-Unit rRNA sequences with curated taxonomy. Nucleic Acids Res. *41*, D597-D604. https://doi.org/10.1093/nar/gks1160.
36. Edgar, R.C. (2010). Search and clustering orders of magnitude faster than BLAST. Bioinformatics *26*, 2460-2461. https://doi.org/10.1093/bioinformatics/btq461.
37. Katoh, K., Misawa, K., Kuma, K., and Miyata, T. (2002). MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. *30*, 3059-3066. https://doi.org/10.1093/nar/gkf436.
38. 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*, 1972-1973. https://doi.org/10.1093/bioinformatics/btp348.
39. Kozlov, A.M., Darriba, D., Flouri, T., Morel, B., and Stamatakis, A. (2019). RAxML-NG: a fast, scalable and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics *35*, 4453-4455. https://doi.org/10.1093/bioinformatics/btz305.

40. Letunic, I., and Bork, P. (2021). Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. *49*, W293-W296. https://doi.org/10.1093/nar/gkab301.
41. Richter, D.J., Berney, C., Strassert, J.F.H., 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 J. *2*, e56. https://doi.org/10.24072/pcjournal.173.
42. Burki, F., Corradi, N., Sierra, R., Pawlowski, J., Meyer, G.R., Abbott, C.L., and Keeling, P.J. (2013) Phylogenomics of the intracellular parasite *Mikrocytos mackini* reveals evidence for a mitosome in Rhizaria. Curr. Biol. *23*, 1541-1547. https://doi.org/10.1016/j.cub.2013.06.033.
43. Minh, B.Q., Schmidt, H.A., Chernomer, O., Schrempf, D., Woodhams, M.D., con Haeseler, A., and Lanfear, R. (2020). IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol. Biol. Evol. *37*, 1530-1534. https://doi.org/10.1093/molbev/msaa015.


A
B



Translation
Transcription
CI
CII
CIII
CIV
CV
Heme
CytC
Transport
total genes
ALVEOLATA
STRAMENOPILES
RHIZARIA
HAPISTA + Telonemids
PROMETHEIDS
ARCHAEPLASTIDA
CRYPTISTA
OPISTHOKONTA
AMOEBOZOA
CRuMS
DISCOBA
Alexandrium
Chromera
Plesmodium
Tetrahymena
Acavomonas
Nyctotherus
Chattonella
Phaeodactylum
Cafeteria
Blastocystis
Bigelowiella
Paracercomonas
Brevimastigomonas
Microkorses
Gephyrocapsa
Chrysochromulina
Marophrys
Nibbleromonas
Ancoracysta
Meteora
Picozoa MS584-11
Cyanidioschyzon
Porphyra
Gleucocystis
Cyanophora
Chlamydomonas
Prototheca
Pycnococcus
Nephroselmis
Mesostigma
Marchantia
Nicotiana
Rhodomonas
Hemiselmis
Palpitomonas
Microhelella
Homo
Allomyces
Saccharomyces
Monosiga
Nuclearia
Thecamonas
Vermamoeba
Acanthamoeba
Dictyostelium
Diphylleia
Mantamonas
Ancyromonas
Malawimonas jac.
Malawimonas cal.
Reclinomonas
Andalucia
Naegleria
Acrasis
Amoebold BB2
Tsukubamonas
Trypanosoma
Euglena gracilis


completeness
0%
100%
alternative positions for telonemids
C60 p-value
UDM64 p-value
rejection
Cryptosporidium muris
Hammondia hammondi
Vitrella brassicaformis
Digyalum owenii
Oxyrrhis marina
Symbiodinium sp.
Oxytricha trifallax
Tetrahymena thermophila
Blepharisma japonicum
unknown Colponemidia
ALVEOLATA
99 | 99
Bolidomonas pacifica
Thalassiosira rotula
Pseudopedinella elastica
Chattonella subsalsa
Paraphysomonas bandaiensis
Saprolegnia diclina
Aplanochytrium kerguelense
Blastocystis hominis
Cafeteria sp.
STRAMENOPILES
91 | 92
91 | 98
.25 .07
Amorphochlora amoebiformis
Orciraptor agilis
Plasmodiophora brassicae
Lapot gusevi
Elphidium margaritaceum
Reticulomyxa filosa
Astrolonche serrata
RHIZARIA
SAR
- | 93
Microkorses curacao
Telonema subtile
Arpakorses idiomastiga
unknown Telonemia
Pavlova sp.
Prymnesium parvum
Choanocystis sp.
Raphidiophrys heterophryoidea
HAPTISTA +
Telonemids
90 | 98
BS%
C60
UDM64
fast-evolving sites removed (%)
.06
Ancoracysta twista
Nebulomonas marisrubri
Nibbleromonas kosolapovi
Ubysseya fretuma
PROVORANS
PROMETHEIDS
Hemimastix kukwesjijk
Spironema cf. multiciliatum
Meteora sporadica
HEMIMASTIGOPHORANS
METEORIDS
89 | 100
Arabidopsis thaliana
Coleochaete scutata
Coccomyxa sp.
Pyramimonas parkeae
Cyanophora paradoxa
Galdieria sulphuraria
Porphyridium purpureum
Picozoa COSAG02
Rhodelphis marinus
ARCHAEPLASTIDA
94 | 91
100 | 98
100 | 98
Chroomonas mesostigmatica
Goniomonas pacifica
Roombia truncata
Palpitomonas bilix
Microheliella maris
CRYPTISTA
Batrachochytrium dendrobatidis
Schizosaccharomyces pombe
Mitosporidium daphniae
Rozella allomycis
Nuclearia sp.
Mus musculus
Trichoplax adhaerens
Stephanoeca diplocostata
Capsaspora owczarzaki
Creolimax fragrantissima
OPISTHOKONTA
91 | -
100 | 97
97 | 92
Amastigomonas sp.
Thecamonas trahens
APUSOMONADS
Lenisia limosa
BREVIATES
- | 93
Pygsuia biforma
Stereomyxa ramosa
Trichosphaerium sp.
Vannella sp.
Vexillifera sp.
Dictyostelium lacteum
Filamoeba nolandi
AMOEBOZOA
100 | 99
AMORPHEA
Diphylleia rotans
Rigifila ramosa
Mantamonas vickermani
CRuMs
96 | -
97 | 92
Fabomonas tropica
Nutomonas longa
ANCYROMONADS
Gefionella okellyi
MALAWIMONADS
93 | -
Eutreptiella gymnastica
Neobodo designis
Naegleria gruberi
Andalucia godoyi
DISCOBA
91 | -
Carpediemonas membranifera
Dysnectes brevis
Monocercomonoides sp.
Trimastix marina
METAMONADA
0.3