Charles University

Faculty of Science

Study programme: Evolutionary Biology

Bc. Marek Valt

Morphological, ultrastructural, and transcriptomic characterization of the newly discovered protist lineage SUM-K

Morfologická, ultraštruktúrna a transkriptomická charakterizácia novo objavenej protistnej línie SUM-K

Diploma thesis

Supervisor: prof. RNDr. Ivan Čepička, Ph.D.

Co-supervisor: Mgr. Tomáš Pánek, Ph.D.

Prague, 2021

Čestné vyhlásenie

Čestne vyhlasujem, že svoju diplomovú prácu som vypracoval samostatne a všetku použitú literatúru a informačné zdroje uvádzam v zozname referencií. Táto práca, ani jej podstatná časť, nebola použitá pre získanie iného alebo rovnakého akademického titulu.

V Prahe, dňa 11. 8. 2021

_____

Marek Valt

Poďakovanie

V prvom rade sa chcem poďakovať celému tímu čepičkalab za vytvorenie skvelej pracovnej atmosféry a priateľského kolektívu. Osobitné ďakujem patrí môjmu školiteľovi, Ivanovi Čepičkovi, a konzultantovi, Tomášovi Pánekovi, ktorí ma za posledné 2 roky naučili viac než som si kedy vedel predstaviť. Rovnako ďakujem aj Billovi Bourlandovi za pomoc s úpravou gramatiky.

Abstrakt

V tejto diplomovej práci predstavujem novoobjavený, hlboko vetviaci organizmus, bičíkovca SUM-K. SUM-K je morfologicky nenápadný protist, ktorý sa vyskytuje v dvoch jedinečných formách, nepohyblivej „slniečkovej“ forme a pohyblivej, bičíkatej forme. Bunky podobné slnku sú guľaté a približne 5 μm veľké, pričom produkujú množstvo tenkých radiálnych výbežkov pripomínajúc Slncovky. Na základe ultraštruktúrnej analýzy a pozorovania pod svetelným mikroskom sme tieto projekcie klasifikovali ako nový typ extruzómov s predpokladanou funkciou predácie koristi. Bičíkaté štádium sa pohybuje jedným predným, posterolaterálne smerujúcim bičíkom. Posteriálny koniec bunky nesie dlhú extrúziu, ktorá pripomína bičík, ale nie je pohyblivá. Na základe týchto morfologických charakteristík nemožno organizmus zaradiť do žiadnej z existujúcich eukaryotických superskupín. Podobne aj fylogenetická analýza založená na SSU rDNA nedokázala vyriešiť evolučnú pozíciu SUM-K, a priradilo ju do príbuzenstva k *Ancoracyste twiste*. Robustná analýza založená na 206 ortologických génoch priradila SUM-K do príbuzenstva so superskupinou Hemimastigophora. Napriek tomu, že morfológia v súčasnosti popísaných hemimastigophoreanov a organizmu SUM-K je úplne rozdielna, ich príbuzenstvo ponúka zaujímavý pohľad na vývoj skupiny Diaphoretickes. Energetický metabolizmus SUM-K je rovnako obzvlášť zaujímavý. Tento protist je rutinne kultivovaný za anoxických a mikrooxických podmienok, ale transkriptomické dáta ukázali, že okrem dráh typických pre aeróbov obsahuje aj bohatú sadu enzýmov anaeróbneho metabolizmu, ako je pyruvát: ferredoxín oxidoreduktáza (PFO), alternatívna oxidáza (AOX) alebo pyruvát:format lyáza (PFL). Bežným motívom medzi anaeróbnymi protistami je strata komponentov elektrón transportného reťazca, čo platí aj pre SUM-K, kde zdanlivo chýba komplex I. Naviac, aj mitochondriálny genóm sa ukázal byť prekvapivo bohatý, obsahujúc vôbec prvý popísaný homológ génu secA, pozostatok ancestrálnej alfaproteobakteriálnej dráhy sekrécie proteínov.

Kľúčové slová: extruzómy, eukaryotická diverzita, elektrón transportný reťazec, anaeróbny metabolizmus, mitogenóm

Abstract

Here I am presenting a novel deep-branching organism, flagellate SUM-K. SUM-K is morphologically inconspicuous protists that occurs in two distinctive forms, immotile "sun-like" form and motile, flagellated form. Sun-like cells are rounded approximately 5 μm big cells that produce numerous thin radial projections artificially resembling Heliozoeans. Based on ultrastructure analysis and light microscopy observations we classified these projections as novel type of extrusomes with putative function of prey predation. Flagellate state moves with one anterior, posterolaterally directed flagellum. The posterior end of the cell bears a long extrusion that resembles a flagellum but is not motile. Based on these morphological characteristics organism can't be classified into any of the establish eukaryotic supergroups. Likewise, phylogenetic analysis based on SSU rDNA failed to resolve its evolutionary position, placing it to relation with *Ancoracysta twista*, an orphan lineage similarly unclassifiable to any of current eukaryotic supergroups. Robust analysis based on 206 orthologous genes placed SUM-K to relation with Hemimastigophora. Even though the morphology of currently described Hemimastigophoreans and organism SUM-K is incongruent, its relation presents interesting prospective of evolution of ancient Diaphoretickes. Energetic metabolism of protist SUM-K is also compelling. We are routinely cultivating it under anoxic and microoxic conditions but transcriptomic data shows a rich set of enzymes of aerobic metabolism in addition to pathways typical for anaerobes, such as pyruvate:ferredoxin oxidoreductase (PFO), alternative oxidase (AOX) or pyruvate formate lyase (PFL). A common theme among anaerobic protists is the serial loss of components of the electron transport chain, which also applies for SUM-K, seemingly absent of complex I. Furthermore, its mitochondrial genome is surprisingly complex, coding first ever described homolog of gen secA, remnant of ancestral alphaproteobacterial pathway of protein secretion.

Key words: extrusomes, eukaryotic diversity, electron transport chain, anaerobic metabolism, mitogenome

Content

1. Introduction and Aims

Protists represent the majority of main eukaryotic evolutionary lineages and constitute a critically important part of microbial communities in a majority of habitats. Since the demise of the Chromalveolata hypothesis (Burki et al., 2007), the eukaryotic tree of life (eToL) has been reworked multiple times. The advent of the "molecular era" allowed discovery of new deep-branching protist lineages with major impacts on our understanding of eukaryotic evolution and supergroup-level diversity. At present, new supergroups are emerging and other groups are being redefined or further resolved. Currently, there are at least 9 eukaryotic supergroups recognized by the majority of protistologists (Adl et al., 2019), generally divided into three "megagroups" – Excavata, Amorphea and Diaphoretickes (Cavalier-Smith, 2002) (Fig. 1).

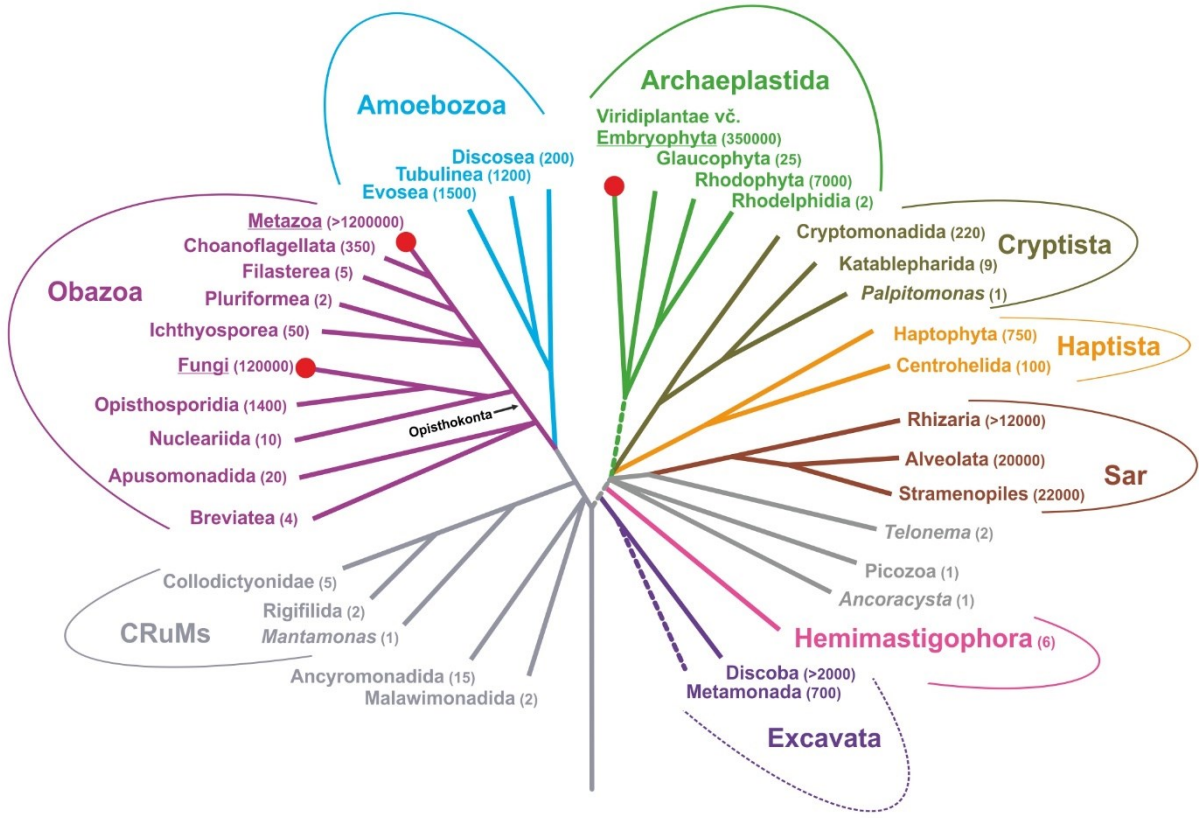

Figure 1: Scheme of the current understanding of the main lineages of eukaryotic organisms based on phylogenomic studies published in the last decade. The numbers in parentheses represent the estimated number of species described. The traditional kingdoms of multicellular organisms - plants, fungi and animals (non-protist eukaryotic lineages) - are marked with red circles. Modified form Čepička (2019).

Two of the newest additions are supergroups CRuMs and Hemimastigophora (Brown et al., 2018; Lax et al., 2018) that have been recognized only during the last three years. Similarly, the supergroup TSAR (Telonemia, Stramenopiles, Alveolata, and Rhizaria) was formed by grouping the well-established SAR group with the phylum Telonemia (Strassert et al., 2019). Internal phylogenetic relationships of the supergroup Amoebozoa were recently updated showing that the ancestor of the group was a flagellate with complex life cycle (Kang et al., 2017). The second supergroup of Amorphea, Obazoa, had its internal phylogenetic relations further improved by discovery of the small flagellates *Pigoraptor* and *Syssomonas* (Hehenberger et al., 2017). The group itself was also created only recently as a successor of the well-known group Opisthokonta (including animals and fungi) to include breviates and apusomonads, all thanks to the discovery of the plesiomorphic flagellate *Pygsuia biforma* (Brown et al., 2013). In the megagroup Diaphoretickes, recent discovery of Rhodelphidia as sister group to red algae, has shown that an ancestor of the whole supergroup Archaeplastida (including land plants) was likely a predatory photosynthetic flagellate (Gawryluk et al., 2019). Finally, the latest placement of group Picozoa inside the supergroup Archaeplastida showed that not all archeaplastids are bound to possess chloroplast (Schön et al., 2021). Discoveries of novel deep-branching protist lineages have also attracted a lot of attention. For example, *Ancoracysta twista* with a gene-rich mitochondrial genome and a unique type of extrusomes, was classified as an orphan lineage that could not be placed into any existing eukaryotic supergroup (Janouškovec et al., 2017).

A third megagroup, Excavata, was formerly used for grouping of Discoba, Metamonada and Malawimonada but it was recently showed to be non-monophyletic, so only informal term “excavates” lingered for its sentimental and, to a point, practical value, as these groups share distinctive morphology, particularly the feeding groove and associated cytoskeleton system (Adl et al., 2019). Malawimonads whose molecular signal suggests a relationship to metamonads, are especially troublesome because most of the recent phylogenomic studies placed them as basal group of Amorphea i.e., on the opposite side of eToL as metamonads (Heiss et al., 2018; Keeling and Burki, 2019; Burki et al., 2020). Therefore, even the truncated excavates, consisting of Discoba and Metamonada, should not be used, as this topology could represent an analysis artefact (Burki et al., 2020).

Phylogenomic placement of the group Malawimonadida also played an important role in the context of rooting the eToL as all analyses using molecular phylogenies agreed that its position lies somewhere near the former group Excavata. Roughly speaking, the eToL

has been divided on Amorphea and Diaphoretickes (Derelle and Lang, 2012; Derelle et al., 2015) or alternatively, between Discoba plus Metamonada and all the other groups (He et al., 2014). While Derelle and Lang (2012) and Derelle et al. (2015) placed malawimonads on the opposite side of the eToL as the rest of the former excavates, He et al. (2014) did not include this group at all. Nevertheless, both these topologies propose that the ancestral properties of the last eukaryotic common ancestor (LECA) had to be rather "excavate-like" as morphological synapomorphies that defined excavates in the first place are present on both branches of the eToL (which also means that the only monophyletic group containing all excavates is whole domain Eukaryota).

If untangling relationships between eukaryotic supergroups and defining the character of LECA, the endpoint of eukaryogenesis, seems challenging enough, the more fundamental task of defining the path taken by protoeukaryotes from the first eukaryotic common ancestor (FECA) to LECA remains in the realm of speculation and controversial hypotheses. Most scientists agree that FECA was chimeric in origin and that one fraction of eukaryotic genes have bacterial ancestry (most notably the alphaproteobacterial ancestor of mitochondria (Betts et al., 2018)), while the rest originated within an archaea, primarily of the group Lokiarchaeota (Spang et al., 2015; Dacks et al., 2016; Zaremba-Niedzwiedzka et al., 2017). Contrary to the simplified belief, the transition between the prokaryotic and eukaryotic cell was not instantaneous and boundaries of this mosaic get blurrier the closer we look. Therefore, FECA and LECA represent the oldest and youngest confines for eukaryogenesis respectively. In between lies a land of uncertainties, complicated by gene duplications and innovations, multiple horizontal gene transfers (HGTs) from numerous bacterial donors, and host-mitochondrion- related endosymbiotic gene transfers (EGTs), i.e., from mitochondrion to host nucleus, all hidden from the eyes of molecular phylogeneticists behind curtains of substitution saturation and a hopeless lack of sampling from extinct protoeukaryotes (Pittis and Gabaldón, 2016; Roger et al., 2021; Vosseberg et al., 2021).

The role of mitochondria in eukaryogenesis has been topic of heated debated. Some argue that this essential endosymbiotic event was the primary spark that ignited eukaryogenesis (Martin et al., 2017), others advocate for the late acquisition of mitochondria (Pittis and Gabaldón, 2016; Hampl et al., 2019; Cavalier-Smith and Chao, 2020). The most recent molecular clock reconstructions present a path of compromise and argue for a "mitochondria-intermediate" hypothesis (Vosseberg et al., 2021). Regardless, there is general agreement that, during eukaryogenesis, the original alphaproteobacterial

endosymbiont genome was significantly reduced and the metabolic pathways were rearranged, which ultimately resulted in the emergence of the modern-day mitochondrial organelle. Even though all mitochondria share a monophyletic origin (Roger et al., 2017; Martin, 2018), they have diversified greatly during the evolution of eukaryotes. The most recurring pattern is miniaturization of mitochondrial genomes to the point where the majority of the previously alphaproteobacteria- encoded genes were either transferred via EGT to the nucleus of the host or were completely lost (Timmis et al., 2004; Nývltová et al., 2015). As a result, the canonical mitochondrial genome encodes only a handful of proteins while the much greater portion of the mitochondrial proteome is encoded in the nucleus and is often of different origin (Gray, 2015).

The oxymonad *Monocercomonoides exilis* took this miniaturization to an extreme level and secondarily lost the mitochondria altogether (Karnkowska et al., 2016). This obligate symbiont is the only known secondarily amitochondriate species (with a possible exception of its close relatives (Hampl, 2017)), showing that presence of mitochondria is not a must for eukaryotic cells and also giving support for the "mitochondria-late" hypothesis (Karnkowska et al., 2019). On the other end of the spectrum there are kinetoplastids with their absurdly DNA-engorged kinetoplast. The estimated mitogenome size of *Diplonema papillatum* (250 Mbp), is the highest yet recorded for any mitochondrial organelle and yet it carries only 18 protein coding-genes (Lukeš et al., 2018). This is due to a highly complex mitochondrial genome architecture which functions mainly as RNA editing machinery (Kaur et al., 2020). An adaptive explanation for this apparent monstrosity remains elusive; but that is not the case for anaerobic protists and their reduction of mitochondrial metabolism. Perhaps the best-known function of mitochondria is its role in ATP biosynthesis via oxidative phosphorylation. Anaerobic protists, in the process of adaptation to life without oxygen, lost many aspects of the ETC (electron transport chain) required for respiration in which oxygen is the electron acceptor, simply because they were not needed and represented only a burden on scarce metabolic and energetic resources (Gawryluk and Stairs, 2021). Today, there is a generally-accepted scheme of mitochondria and mitochondrion-related organelle (MRO) classification, based on their different types of energy metabolism. Aerobic mitochondria are assigned as class I, anaerobically functioning mitochondria as class II, hydrogen-producing mitochondria as class III, hydrogenosomes as class IV, and the most reduced MROs, mitosomes, as class V (Müller et al., 2012). Although, this system has practical value, it is deceptively simplistic. First, most of the novel protist lineages tend to

defy this classification and their mitochondrial metabolism appears to be somewhere on the spectrum in between these classes (Stairs et al., 2015; Leger et al., 2017; Roger et al., 2017; Gawryluk and Stairs, 2021). We create categories but nature selects blindly, giving no regards. Thereby, it would be counterproductive to assign each and every novel and slightly diverse mitochondrion to its own class. Second, canonical aerobic mitochondria (class I), as seen in conventional model organisms, are metabolically rather uniform organelles. From this classification it would seem that they represent the "apex" of evolutionary adaptation, but this is certainly not the case. All attempts to reconstruct proto-mitochondrial metabolism lead to the conclusion that the mitochondrion of LECA had remarkably complex abilities and was capable of both aerobic and anaerobic respiration (Gabaldón and Huynen, 2003; Koumandou et al., 2013). In recent years, characterization of distant anaerobic protist lineages further supported the hypothesis of bacterial-like proto-mitochondrial metabolism, e.g., discovery of a bacterial type II secretion system in malawimonads, jakobids, heteroloboseans, and hemimastigophoreans (Horváthová et al., 2021), or the mitochondrial peptide-based protein-targeting system inherited from LECA and homologous to the bacterial signal recognition particle (SRP) system (Pyrih et al., 2021). Overall, a clearer pattern is emerging, showing that the plesiomorphic state of mitochondria was clearly more bacterial-like then we previously assumed. Record holders in these regards are jakobids *Reclinomonas americana* and *Andalucia godoyi*, the latter with exactly 100 mitochondrially encoded genes, representing the most gene-rich mitochondrial genome known to date (Burger et al., 2013) and 865 nucleus-encoded mitochondrial proteins (Gray et al., 2020; Valach et al., 2021).

Organism SUM-K was first discovered in our laboratory as a eukaryotic contaminant during metagenome assembly from an unrelated project. This highly fragmented and contaminated metagenomic data gave us the first steppingstone from which we could advance towards expanding our knowledge of this organism.

The single aim of this diploma thesis is to describe organism SUM-K as comprehensively as possible, given the data acquired to date. Since its initial recognition by light-microscopy, we knew we were dealing with something extraordinary. Study of its ultrastructure extended this knowledge and revealed an entirely new type of extrusome. Phylogenomic analysis of the SSU rRNA gene hinted at a deep evolutionary position that was later validated by a transcriptomic approach. Based on this data we could also begin exploration of expressed genes to uncover the nature of its metabolism, distinguishing traits,

and ancestral metabolic pathways that are crucially important in understanding the evolutionary origins of eukaryotes. SUM-K provides the rare opportunity to carry out these investigations in one single flagellate. Even now it seems we are just "scratching the surface" with regard to the unique features of this organism. At every turn in our studies of it, we found something noteworthy, exceptional, or even textbook-changing. Some of our findings are presented in the following pages.

2. Material and methods

2.1. Sampling and cultivation

The original sample from which we established a culture of organism SUM-K was obtained from a marine sediment (depth 30 m) on the coast of a Croatian's island Brač, on the outskirts of the city of Summartin (- 43°17' N 16° 52' E) in 2011. A monoeukaryotic culture of SUM-K is currently being grown in polyxenic culture with unidentified bacteria at room temperature in American Type Culture Collection (ATCC) Seawater Cereal Grass Media #1525. Cultures are maintained by weekly reinoculation of approximately 1 ml of the culture sediment into a 15 ml tube with approximately 10 ml of fresh media. Cultivation is ongoing in both microoxic and anoxic conditions. A microoxic environment is sustained by cultivation in firmly tightened culture tubes where bacteria present in the culture consume most of the available oxygen, achieving an optimal microoxic equilibrium. Anoxic cultivation takes place in an anerobic chamber with an anaerobic atmosphere generation system (AnaeroGen 2.5, Thermo Fisher Scientific).

2.2. Light microscopy

An Olympus BX51 compound microscope equipped with differential interference contrast (DIC) and brightfield (BF) illumination was used for light microscopy. Video recordings were made on NIKON TiE 2, wide-field, inverted microscope equipped with DIC and monochromatic camera capable of video capture up to 100 frames per second (fps). Subsequent analysis of high fps data was made on NIS-Elements AR 5.21.00 software.

2.3. Protargol staining

For protargol staining, cells were fixed according to the protocol by Bodian (1936) and later modified by Nie (1950), which was further optimized by our laboratory group members for

staining of small flagellates as follows: On a coverslip, 1 µm of culture sediment was mixed with drop of a sterile egg albumin. To achieve better adherence of the cells egg albumin was not diluted. The coverslips were then fixed in Bouin-Hollande's solution (0,175 mol/l picric acid, 0,138 mol/l Cu^{+2} acetate, 4% formaldehyde, and 5% acetic acid in water solution) over night. On the next day, the coverslips were transferred through a graded ethanol series (70%, 50%) and washed in distilled water, then bleached in a 0,5% $KMnO_4$ solution for 5 minutes. After bleaching, the coverslips were washed in distilled water (5x), treated with 5% oxalic acid for 5 minutes and washed again in distilled water (5x) before staining in 1% protargol solution (Bayer, I. G. Farbenindustrie Actinengesellschaft; out of business since 1952). Staining lasted for 48 hours at the temperature 37°C in a beaker with copper wire pieces placed in between the coverslips. After protargol staining, the coverslips were washed in distilled water (2x), treated with freshly prepared reducing solution (1% hydroquinone and 5% Na_2SO_3) for 10 minutes, washed again in distilled water (5x), and toned with 1% $AuCl_3$ for 5 minutes. Next, the washed coverslips (2x) were treated with 2% oxalic acid for 5 minutes, washed in distilled water (5x), and treated with 5% $Na_2S_2O_3$ for 10 minutes. Final washing was done under constant stream of tap water for 20 minutes. The coverslips were then dehydrated in an ethanol series (50%, 70%, 80%, 96%, 100%) and 3 times in xylene (5 minutes each). Finally, the stained, dehydrated coverslips were mounted on glass slides with DPX mounting medium (Sigma – Aldrich). Images were captured on Olympus BX51 microscope with BF illumination.

2.4. Transmission electron microscopy (TEM)

To study ultrastructure, cells were fixed with 2,5% glutaraldehyde fixative and cacodylate buffer solution and postfixed with 1% osmium tetroxide. The fixed samples were next washed in distilled water, dehydrated in ethanol series (30%, 50%, 70%, 80%, 90%, 95%, 100% x3) and impregnated with acetone 1:1 ethanol, 100% acetone and acetone 1:1 EPON-araldite resin mixture. Afterwards, samples were transferred to EPON resin (Poly/Bed 812, Polysciences) and polymerized for 48h at 70°C. Further processing of samples, including preparation of serial ultrathin (80 nm) sections, mounting on grids, and staining with lead citrate and uranyl acetate were carried out as a service of the Laboratory of Electron Microscopy, Faculty of Science, Charles University. Finally, prepared sections were examined with JEOL 1011 transmission electron microscope (JEOL LV) with a Veleta CCD camera, utilizing Olympus Soft Imaging Solution software.

2.5. DNA isolation, amplification, cloning and sequencing of the SSU rRNA gene

Total DNA of SUM-K was isolated using the Genomic DNA Mini Kit (Geneaid) according to the manufacturer's instructions. Attempts to amplify the 18S ribosomal rRNA gene (SSU rRNA) using general eukaryotic primers MedA (CTGGTTGATCCTGCCAG) and MedB (TGATCCTTCTGCAGGTTCACCTAC) (Medlin et al., 1988) yielded insufficient products for direct sequencing (Supp. fig. 1). Therefore, the PCR product was cloned using the pGEM-T Easy Vector System I kit (Promega) and JM109 competent cells of *Escherichia coli* (High Efficiency Competent Cells, Promega). The Sanger sequencing of cloning products took place in the DNA Sequence Laboratory at the Faculty of Science, Charles University with an ABI PRISM 3100 sequencer (ThermoFisher Scientific).

2.6. Phylogenetic analysis of the SSU rRNA gene and phylogenomic analysis

For the phylogenetic analysis of the SSU rRNA gene, sequences were aligned using the MAFFT online service (Katoh et al., 2019) utilizing L-INS-i algorithm and then manually edited and trimmed in BioEdit software (Hall, 2004). Phylogenetic reconstruction was preformed using maximum likelihood (ML) method in RAxML (Stamatakis, 2006) with model GTRGAMMAI and statistical topology support was estimated by non-parametric bootstraps with 100 replicates.

Pan-eukaryotic phylogenomic analysis, from transcriptomic data, was performed by Dr. Matthew W. Brown (Mississippi State University, USA), a collaborator of my supervisor and co-supervisor. Phylogenomic tree of 252 eukaryotes was inferred from 206 orthologs using ML method with the LG+Γ4+C60-PMSF substitution model, and an LG+ Γ4+C20 ML tree as a posterior mean site frequency (PMSF) guide input tree in software IQ-TREE v2.0-rc1 (Minh et al., 2020). Maximum Likelihood bootstraps (MLBS) were inferred from collected PMSFs from the above-described ML analysis and used for non-parametric bootstraps with 100 replicates. The final supermatrix of 61995 AA (aminoacid) sites as well as single-protein alignments and subsequent supermatrix-based phylogenomic analysis were processed with the PhyloFisher software with settings as recommended by Tice et al. (2021).

2.7. RNA isolation and Illumina transcriptome sequencing

Total RNA was extracted from 150 ml of well-grown monoeukaryotic culture by the method described in the TRIzol-chloroform RNA isolation protocol (Rio et al., 2010) using TriReagent (Sigma – Aldrich), then further purified by RNeasy mini kit (Qiagen) and treated with DNase (Qiagen) for elimination of contaminating DNA. Isolated and purified RNA was stored at -80 °C. Quality assessment was done with agarose gel electrophoresis (Suppl. fig. 2) and a 2000 UV-Vis spectrophotometer NanoDrop at the DNA Sequence Laboratory (Faculty of Science, Charles University). The measured concentration was 725,9 ng/µl with wavelength ratios 2,21 (260/280) and 2,24 (260/230) (Suppl. Fig. 2). Purified RNA was sent to Macrogen Europe (Amsterdam, NL). There, poly-A mRNA selection, an additional quality control, and cDNA library construction with random fragmentation was made, using the TruSeq stranded mRNA library kit. Sequencing was performed on the Illumina NovaSeq 6000 platform for 2 disparate samples in order to minimize sequencing errors. A summary of quality assessment and statistics of raw generated data is shown in Supplementary table 1.

2.8. Assembly, construction, and cleaning of prokaryotic contaminants

For construction of the transcriptome assembly, we first cleaned raw RNA-seq data of adapters and poor quality readings using the Trimmomatic software (Bolger et al., 2014). The data from two separate sequencing runs were then assembled with Trinity software (Grabherr et al., 2011) and clustered in software CD-HIT (Fu et al., 2012) with a 100% sequence identity cut-off. Because SUM-K is being grown in polyxenic culture with unidentified bacteria, the data were significantly contaminated with prokaryotic sequences.

For achieving the best final assembly quality possible, extensive cleaning of a prokaryotic contaminants was done. Metagenomic sequences that were acquired during an unrelated study in our laboratory, in which SUM-K was, itself, a eukaryotic contaminant, were used for identification of the prokaryotic organisms putatively inhabiting the polyxenic culture (metagenome was sequenced using Illumina NextSeq as described in (Beinart et al., 2018). All metagenomic contigs longer than 100 kbp (kilobase pairs) were selected and verified to be of prokaryotic origin. Verification was accomplished by prediction of several proteins longer than 200 AA within the frame of 8000 bp long fragments of each contig. Subsequently, we used the blastp module of BLAST (Agarwala et al., 2016) against a local copy of the NCBI protein database on these predicted proteins. A maximum of 10 best hits for each protein with e-value lower than $1e^{-100}$ were retained and each hit was assigned with

affiliation to its domain (Archaea, Bacteria, Eukarya, or Viruses). Potentially eukaryotic contigs were then manually deleted and sequences were archived for later steps. A second source of identification of prokaryotic contaminants were 16S rRNA gene sequences within transcriptome data. Identified sequences were cleared of any obvious chimeras and used as query against the aforementioned metagenomic scaffolds with the blastn module of BLAST (Gerts et al., 2006). The longest alignments for each query were manually selected and refined so only hits longer than 1000 bp and shorter than 100 kbp (already detected in first step) and simultaneously containing the identified 16S rRNA genes from transcriptomic data were retained. All sequences, selected from both steps, were manually confirmed to be of prokaryotic origin so that we would not accidentally include mitochondrial SSU rRNA genes in the final cleaning of the database. The 265 predicted metagenomic 16S rRNA gene sequences were blasted with the blastn module against a local copy of the NCBI nucleotide database. From the results we handpicked closest relatives, with at least 99% identity, with a publicly available genome assembly. If a published genome was fragmented into numerous contigs, we selected only those containing 16S rRNA gene sequences to avoid possible misannotations in NCBI database. After the final dataset of genomes of putative prokaryotic contaminations was ready (list of used bacterial genomes is available at Supplementary table 2), we began with the cleaning itself. For this purpose we used program HISAT2 (Kim et al., 2019) to map RNA reads on selected prokaryotic genomes. Unmapped reads were reassembled into contigs using Trinity (Grabherr et al., 2011) and rnaSPAdes (Bushmanova et al., 2019) software packages. This rather exhaustive procedure allowed us to acquire a transcriptome assembly of extraordinarily high quality with a minimum of prokaryotic contaminants. Based on an analysis using the BUSCOv4 benchmark scoring tool (Manni et al., 2021) with the whole eukaryote lineage dataset, the more complete of the two assemblies was selected – in this case the Trinity assembly (Suppl. Fig. 3).

At last, the contigs of the final Trinity assembly were translated into AA sequences using the TransDecoder software that identifies candidate coding regions within transcript sequences (https://transdecoder.github.io).

Investigation of potential eukaryotic contaminants was carried out on a set of highly-expressed genes, the sequences of which were searched using HMM (hidden Markov models) profiles made by HMMER v3.3.2 (Finn et al., 2011; http://hmmer.org/), followed by phylogenetic analysis using ML with software RAxML (Stamatakis, 2006). After manual inspection, no eukaryotic contaminants were present.

2.9. Automatic annotation

Automatic annotation of the mitochondrial genome was performed with utility MFannot (developed by the labs of B. F. Lang and G. Burger; (unpublished); available at megasun.bch.umontreal.ca/cgi-bin/mfannot/mfannotInterface.pl). Graphical illustrations were made with the OrganellarGenomeDRAW toolkit (Greiner et al., 2019). This analysis was performed by my co-supervisor, Tomáš Pánek and data used were from metagenomic assembly in which SUM-K was originally discovered.

As SUM-K is a non-model, deep-branching protist, we used a triply-redundant automatic annotation approach on transcriptomic data in an attempt to achieve as comprehensive predictions as possible. BlastKOALA (Kanehisa et al., 2016) uses BLAST searches against the KEGG (Kyoto Encyclopedia of Genes and Genomes) database of representative genomes. KofamKOALA (Aramaki et al., 2020) takes a different approach and searches sequence data by HMMSEARCH against a customized HMM database of KEGG orthologous groups. Both these tools were employed remotely with default settings on a database of both eukaryotes and prokaryotes and with e-value cut-off threshold 0,01 for KofamKOALA. The third prediction tool used was eggNOG-mapper v2 (Cantalapiedra et al., 2021) on a locally constructed pan-eukaryotic database inferred from the eggNOG 5.0 database (Huerta-Cepas et al., 2019) with diamond search option that uses DIAMOND sequence aligner designed for high performance analysis of big sequence data (Buchfink et al., 2021).

Before inspection of annotated KEGG results of both approaches, we first merged their outputs and removed duplicate hits. For visualization itself, we used the web service KEGG mapper v5.0 (Kanehisa and Sato, 2020). Separate searches were also observed and compared independently. The KofamKOALA search provided more results of mostly short and fragmented sequences but also with more false positive hits. Complete and more highly-conserved sequences were identified by all of the approaches, differences were mostly in false positive hits and fragments.

2.10. Manual annotations

Identification of orthologs of some enzymes of anaerobic metabolism and other proteins of interest was performed manually because the identification of these proteins and their differentiation from bacterial homologues is not reliable using automated methods (Nguyen et al., 2019). For this we created custom HMM profiles of selected enzymes with the module

hmmbuild, and with module hmmscan we used them to search our transcriptomic database (both modules part of HMMER v3.3.2 (Finn et al., 2011; http://hmmer.org/)). Datasets used for HMM profile construction were created by members of the research team in other research projects or downloaded from PANTHER (Thomas et al., 2003) or Pfam (Mistry et al., 2021) databases. For handling of outputs we used the blastdbcmd BLAST+ module (Camacho et al., 2009) and module SearchIO from the BioPython package (Cock et al., 2009).

2.11. Curation of annotated sequences

Every protein included in this thesis was manually curated and full outputs of automatic annotation are not disclosed. For identification of prokaryotic contaminants that eluded our cleaning strategy, we used a BLAST search and if results were inconclusive, we continued with single gene phylogenetic analysis on a custom-made dataset. Datasets were created with best hits from BLAST searches and representative sequences from all three domains of life (Archaea, Bacteria and Eukarya). We used entries from numerous databases, mainly UniProtKB (Bateman et al., 2021), NCBI RefSeq (O'Leary et al., 2016), or TIGRFAMs and PANTHER seed alignments of trusted representative sequences (Haft et al., 2013; Thomas et al., 2003). Sequences were aligned with multiple sequence alignment software MAFFT v7.453 (Katoh and Standley, 2013) utilizing the L-INS-i algorithm. Alignments were trimmed either manually or with program trimAL (Capella-Gutiérrez et al., 2009). Alternatively, we used trimming with BMGE software (Criscuolo and Gribaldo, 2010) which is designed for minimizing phylogeny reconstruction artefacts. The latter approach was used for computationally intensive inferences. Quick tree inference was performed using FastTree 2 with default parameters (Price et al., 2010) and if needed, a more comprehensive phylogenetic tree was computed with IQ-TREE v. 2.1.2 (Minh et al., 2020) with ultrafast bootstrap approximation (Hoang et al., 2018) and automatic model finder (Kalyaanamoorthy et al., 2017).

For more detailed analysis of selected proteins we used InterProScan 5 (Jones et al., 2014) that classifies proteins into families and predicts conserved domains using signature predictive models. Similarly, we used protein sequence analysis provided by the MPI Bioinformatics Toolkit (Gabler et al., 2020). Primarily, we applied an HHpred search that is based on the pairwise comparison of HMM profiles and is very sensitive in detecting remote homologs.

In an effort to predict mitochondrial targeting presequences, we used various automated tools, including TargetP 1.1 (Emanuelsson et al., 2000), WoLF PSORT (Horton et al., 2007), Mitoprot II (Claros and Vincens, 1996), TPpred 3.0 (Savojardo et al., 2015), MitoFates (Fukasawa et al., 2015), TargetP 2.0 (Armenteros et al., 2019), and DeepMito (Savojardo et al., 2020).

All of the computationally demanding tasks were computed on the MetaCentrum grid that uses computational resources provided by the ELIXIR-CZ project (LM2018131), part of the international ELIXIR infrastructure.

3. Results and discussion

3.1. Light-microscopic morphology

In culture, the organism SUM-K occurs in two distinctive forms, an immotile form – here referred as the "sun-like" form, and a motile, flagellated form. The sun-like cells are rounded and produce numerous thin radial projections, identified as novel type of extrusomes (Fig. 2 A-C) (see 3.2. Ultrastructure). The mean size of the cells is 3,8 μm excluding the radial projections (2,7 μm min./5,2 μm max. – average of 30 measured cells), and the projections can be up to 7,5 μm in length. The shape of the motile flagellated form is elongated ellipsoidal with a single, posterolaterally directed flagellum inserted at the anterior end of the cell and a thick, immotile posterior projection (Fig. 2 D-F). The mean length of the flagellum is 11,3 μm, the length of the posterior projection is 5,3 μm, with length of the cell body 7,9 μm and diameter 3,9 μm (average from 2 clearly-imaged cells). The flagellates were occasionally seen to adhere to the glass slide, absorb the flagellum and the posterior projection, and then become spherical (Fig. 2 G, H). A transitional stage is immotile and looks like a simple spheroid cell (Fig. 2 I). Subsequently, they start to produce radial projections tipped with extrusomes and in the span of several minutes, they become fully-grown sun-like forms. Transformation from sun-like form or transitional spheroid form to flagellated form was not observed. This may reflect the relative numbers of different stages in culture where the flagellated stage is much less abundant than the sun-like form. An exact ratio of their respective abundance is hard to determine, because sun-like cells tend to aggregate in sediment, so their numbers are easily underestimated.

SUM-K is putatively a facultative aerobe. It is being routinely cultured under microoxic and anoxic conditions in the presence of unidentified prokaryotes, which quickly

consume all available oxygen soon after subculturing. Every attempt at aerobic cultivation has failed so far but different culturing methods are being investigated to determine if growth can occur under aerobic conditions.

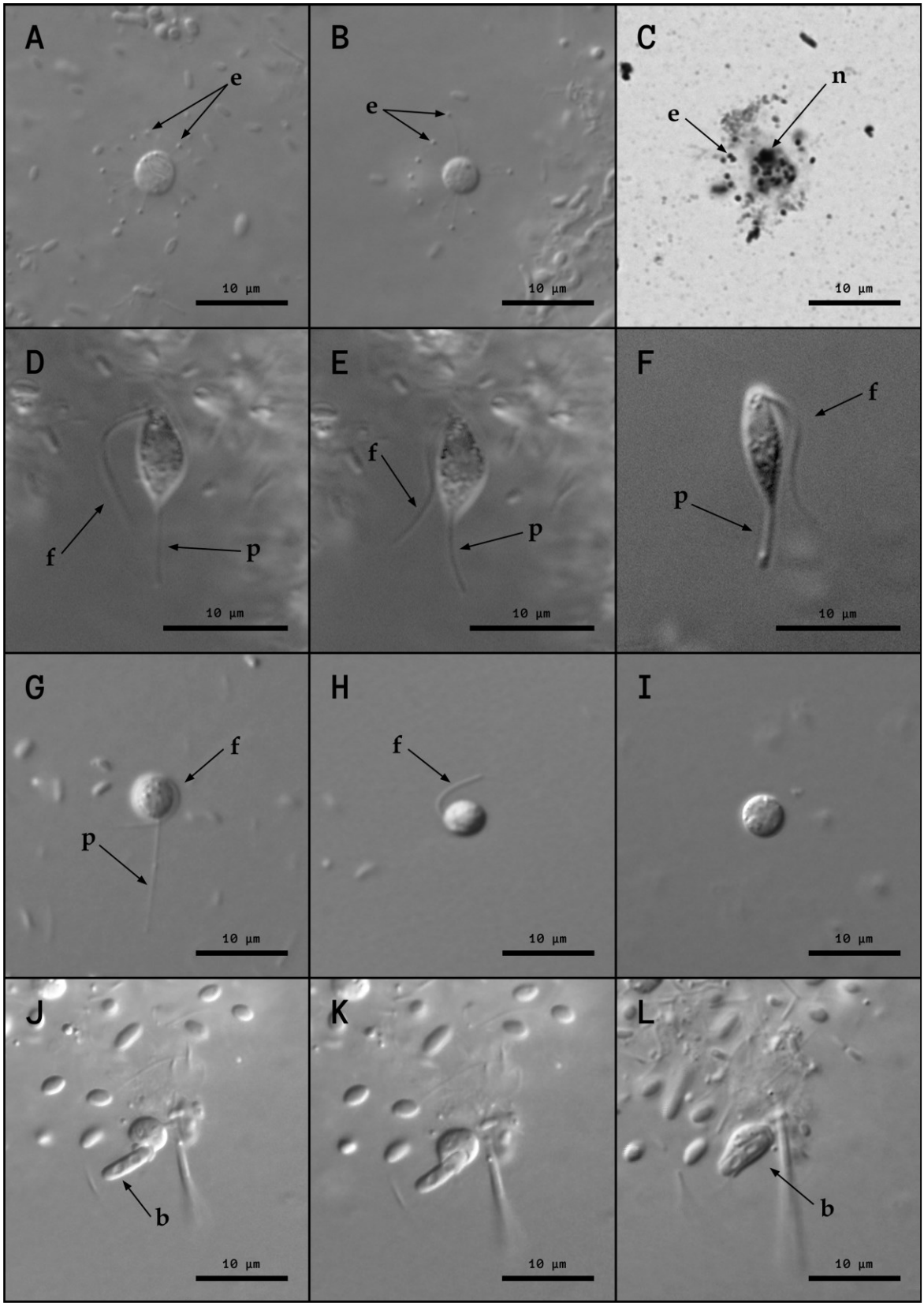

Figure 2: The morphology of the organism SUM-K. **A** and **B**: sun-like forms with extruded extrusomes tipped with globular points. **C**: protargol-stained preparation of sun-like form, with visible nucleus and distinguishing refringent tips of extrusomes. **D-F**: flagellated forms with immotile projection at the posterior end and posterolaterally directed flagellum, inserted in the anterior part. Displayed in different phases of the sinus wave pattern motion. **G-I**: transition from flagellated form to sun-like form. In G, the flagellum starts to encompass the cell but remains motile at the terminal part while posterior projection is getting integrated into the cell body that is changing shape from elongated-teardrop to rounded-globular. In H, the flagellum movement is slowing and eventually stopping altogether, posterior projection is fully absorbed while now fully rounded cell keeps wobbling in one place. In I, flagellum is fully absorbed into the cell, cell is immotile and in the span of several minutes starts extruding extrusomes. **J-L**: depiction of bacterial prey capture. The bacterium, longer than the cell itself, is captured, putatively with assistance of the extrusomes, enveloped by pseudopodia of a phagosome and eventually fully phagocytized. C captured with BF, all the others with DIC contrast. D-F were captured on NIKON TiE2 microscope, rest of the pictures captured on Olympus BX51 microscope. (b – bacterial prey; e – extrusome; f – flagellum; n – nucleus; p – posterior projection).

Notably, the morphologic features of both forms are quite unusual and SUM-K cannot be identified as a member of any known protist lineage based on its morphology. The sun-like form is reminiscent of heliozoans and suctorian ciliates, but the similarity is clearly only superficial and does not indicate a phylogenetic relationship. Sun-like forms are active stages capable of pray capture (Fig. 2 J-L) and they were observed to divide. Movement of the flagellated stage is also remarkable (Suppl. video). The flagellates of SUM-K are unikonts with a posterolaterally directed flagellum that beats posteriorly with a sine wave pattern. The cells revolve on the long axis and tend to bore into organic sediment. This arrangement is relatively rare as unikonts usually have the single flagellum directed anteriorly (e.g., *Phalansterium solitarium* (Ekelund, 2002) Mastigamoeba invertens (Walker et al., 2006), *Creneis carolina* (Pánek et al., 2014)). A similar flagellar arrangement can be observed in an unrelated fornicate *Iotanema spirale* (Yubuki et al., 2017) or plesiomorphic cercozoan *Allantion parvum* (Howe et al., 2011).

Protargol-stained cells show argyrophilic tips of extrusomes, inside and outside of the cell, and an eccentrically-positioned nucleus (Fig. 2 C). Protargol staining of this organism, proved to be quite difficult (from 16 stained coverslips we were able to recover only 3 adequately stained cells, all of which were sun-like stages). This is most probably due

to character of the sediment where a majority of the cells reside. Sediment containing the cells tends to clump together and is often lost form coverslips during washing steps.

3.2. Ultrastructure

One of the key characteristic features of each eukaryotic lineage is the structure of their flagellar apparatus, i.e., the organization of flagella and their associated microtubular and non-microtubular cytoskeletal elements. Only in recent years has it been proved that individual elements of the flagellar apparatus can be homologized across the eToL and, thus, the structure of the flagellar apparatus of the LECA can be reconstructed (Yubuki and Leander, 2013; Yubuki et al., 2016; Adl et al., 2019). We predict that SUM-K will also have a unique structure of the flagellar apparatus. Judging by the direction of the SUM-K flagellum we could hypothesize that it is homologous to the proposed recurrent flagellum of LECA but, as the example of the superficially similar flagellum of *Iotanema* shows, it could actually be an anterior flagellum (Yubuki et al., 2017). Therefore, reliable determination can be only achieved by inspection of the associated cytoskeletal roots. Unfortunately, we were not able to recover any sections of the flagellated stage in our TEM analysis, presumably because of their low abundance in culture. On the other hand, we were able to acquire numerous micrographs of the sun-like stage, 67 individual cells, 39 of which were on more than 2 sections (5 cells in peak of 7 sections).

Transmission electron micrographs of SUM-K showed that sun-like stage cells have everything that a "proper" eukaryotic cell should have (Fig. 3). There is a big, acentric nucleus, sometimes with a sizeable electrondense nucleolus (Fig. 3 A, B). Typically, the nucleus is surrounded by a well-formed endoplasmic reticulum, which can sometimes extend next to the mitochondrion (Fig. 3 A, C, F). Large food vacuoles (Fig. 3 A) frequently occupy the majority of the cell volume. There are often multiple peroxisome-like bodies in one cell (Fig. 3 A, C) and the Golgi apparatus tends to have typical structure (Fig. 3 D). On the external surface of the cell membrane there is a clearly visible glycocalyx (Fig. 3 A). A few cells were, fortuitously, imaged during prey capture and in the process of phagosome (food vacuole) formation (Fig. 3 A, B). Also cells were imaged during karyokinesis, slightly before nuclear division judged by the contractile ring like structure seemingly splitting nucleus (Fig. 3 E). Regarding the mitochondria; they were often numerous, always close to the cell membrane, and seemingly acristate, although the matrix is not filled homogeneously (Fig. 3 A, C, F). It must be mentioned that we were never able to distinguish 2 membranes, essential

for certain identification of mitochondria in TEM sections. Therefore, the organelle resembling a mitochondrion in our TEM analysis is only putative. The only element that is continually missing are microtubular structures except for a few sections including a centriole (Fig. 3, 4 D). Supportive structures that are formed during the phagosome creation are most likely just filamentous structures (Fig. 3 B), the same applies for extrusome projections (see Fig. 4).

In microsection F (Fig. 3 F) we can also see small vesicle exiting mitochondrion. This might be just an artefact created during the fixation process or also might not, if we take into consideration that similar vesicles were observed numerous times in multiple cells (see chapter 3.4. for possible explanation).

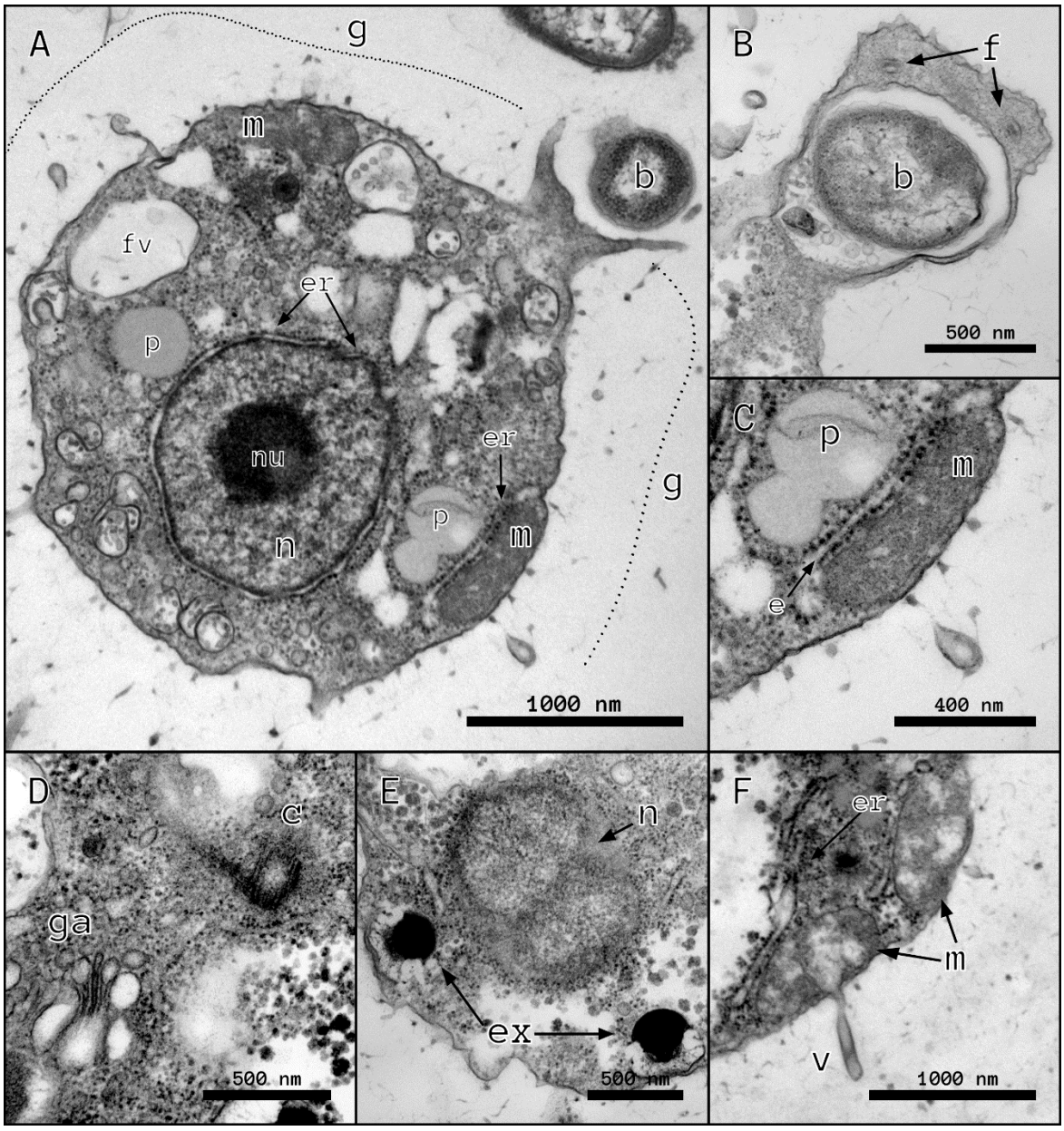

Figure 3: TEM micrographs of sectioned sun-like cells of organism SUM-K. **A**: representative sun-like cell, with dense nucleolus inside large nucleus that surrounded by endoplasmic reticulum extending next to a mitochondrion near the cell membrane. Also shown are multiple peroxisome-like bodies, food vacuole and glycocalyx surrounding the entire cell surface. The cell is pictured in act of prey capture, forming phagosome around bacteria. **B**: detailed image of bacterial prey capture from another cell with partly visible supporting microfilaments. **C**: detail of mitochondrion and peroxisome-like body of the same cell and microsection as A, positioned next to the cell membrane. Endoplasmic reticulum extends near the mitochondrion. **D**: detail of Golgi apparatus and centriole. **E**: detail of nucleus during karyokinesis, with unextruded extrusomes positioned close to the cell membrane. **F**: detail of two mitochondria positioned close to the cell membrane and also endoplasmic reticulum and vesicle, putatively exiting mitochondrion. (b – bacterial prey; c – centriole; er – endoplasmic reticulum; ex – extrusome body; f – microfilaments; fv – food vacuole; g – glycocalyx; ga – Golgi apparatus; m – mitochondrion; n – nucleus; nu – nucleolus; p – peroxisome; v – vesicle).

Only a few of the established eukaryotic supergroups have defining morphological characteristics and, more often than not, they can be categorized only through phylogenomic analysis. Lineage-specific cell organelles are therefore rare and they often have homologous or analogous counterparts on different branches of eToL (Adl et al., 2019). Such organelles include extrusomes i.e., organelles that extend or eject from the cell through the cytoplasmic membrane in response to various stimuli (Hausmann, 1978). Extrusomes have diverse functions, structures, and complexity. Their nomenclature can be referred to as chaotic, with the last comprehensive reviews being decades old and their postulated categorization being followed only loosely (Hausmann, 1978; Kugrens et al., 1994; Rosati and Modeo, 2003). Consequently, novel types of extrusomes are often given their own category (e.g., chromerosomes of the alveolate *Chromera velia* (Oborník et al., 2011), telonemosome of *Telonema subtilis* (Yabuki et al., 2013), vlimatocysts of the dinoflagellate *Oxyrrhis marina* (Rhiel, 2017), and ancoracysts of *Ancoracysta twista* (Janouškovec et al., 2017)).

Here I present such a novel type of extrusome from the organism SUM-K (Fig. 4). I have no ambition to give these extrusomes their own name yet (even though few terms have been discussed, e.g., heliosomes – “sun bodies”). Main focus was at describing their morphology, function, assembly, and interactions with prey and the environment. Each SUM-K cell produces numerous extrusomes (up to about 30), but the exact count is hard to determine because they are produced in several focal planes. Typically, extrusome bodies are abundant under the surface of the cell membrane, but they can be present even in more

central parts of the cell (Fig. 4 A, F). Extrusome projections are produced radially in all directions and extrusome bodies do not change size nor shape after extrusion (Fig. 4 A, B, C). Their function is very likely offensive i.e., predation of prokaryotic prey, but it is possible that they also aid in attachment to debris in its sedentary way of life. Regardless, these functions are not mutually exclusive. Microtubular supportive structures are not clearly visible in extrusome projections, putatively they are supported only by microfilaments. This conclusion is based on observations of over 200 microsections in which we were unable to identify any microtubular structures other than the centriole (Fig. 3, 4 D). Assembly of the extrusome bodies appears to occur at the Golgi apparatus (Fig. 4 D, E) with subsequent maturation in the cytoplasm (Fig. 4 A). This is in concurrence with descriptions of the formation of numerous other extrusomes (Thomsen et al., 1991; Zhou et al., 2011; Plattner, 2017; Buonanno and Ortenzi, 2021). The ultrastructure of extrusome bodies appears to be quite complex. In longitudinal section they are round with a distinct inner circle of higher electron density and they are surrounded by an inhomogeneous matrix, altogether enveloped by a membrane (Fig. 4 B, G, I). On cross-section, more unique features emerge. Firstly, the extrusomes appear hollow, with a shaft traversing the middle part of the body (Fig. 4 H, J, K, N). Possibly, the shaft acts as a connection channel through which attachment with prokaryotic pray is established (Fig. 4 K), but details about the mechanism of prey capture are, as yet, unknown. Different cross-sections show that the internal conformation of the extrusome is not homogenous i.e., and an extrusome cap and main extrusome body can be distinguished (Fig. 4 F, H, L, M). the extrusome body sometimes assumes strikingly geometrically symmetric formations (Fig. 4 N). In serial sections of the same cell (Fig. 4 L, M), we can see the extrusome cap separated from the extrusome body. These two parts are commonly both closely connected to the food vacuole. It is unclear yet if this is common behavior of SUM-K extrusomes and if it is connected to the way in which extrusomes might capture bacterial prey.

For the purpose of identifying similar or identical extrusomes that have been already described we did an extensive search of the literature concerning the ultrastructure of extrusomes or extrusive bodies. The most similar, but nevertheless structurally distinct, extrusomes described are those of *Commation eposianum* (Stramenopiles *incertae sedis*), which differ in having a clearly visible cavity inside the extrusome bodies (Thomsen and Larsen, 1993). Extrusomes of the jakobid *Histiona aroides* are poorly described, but from

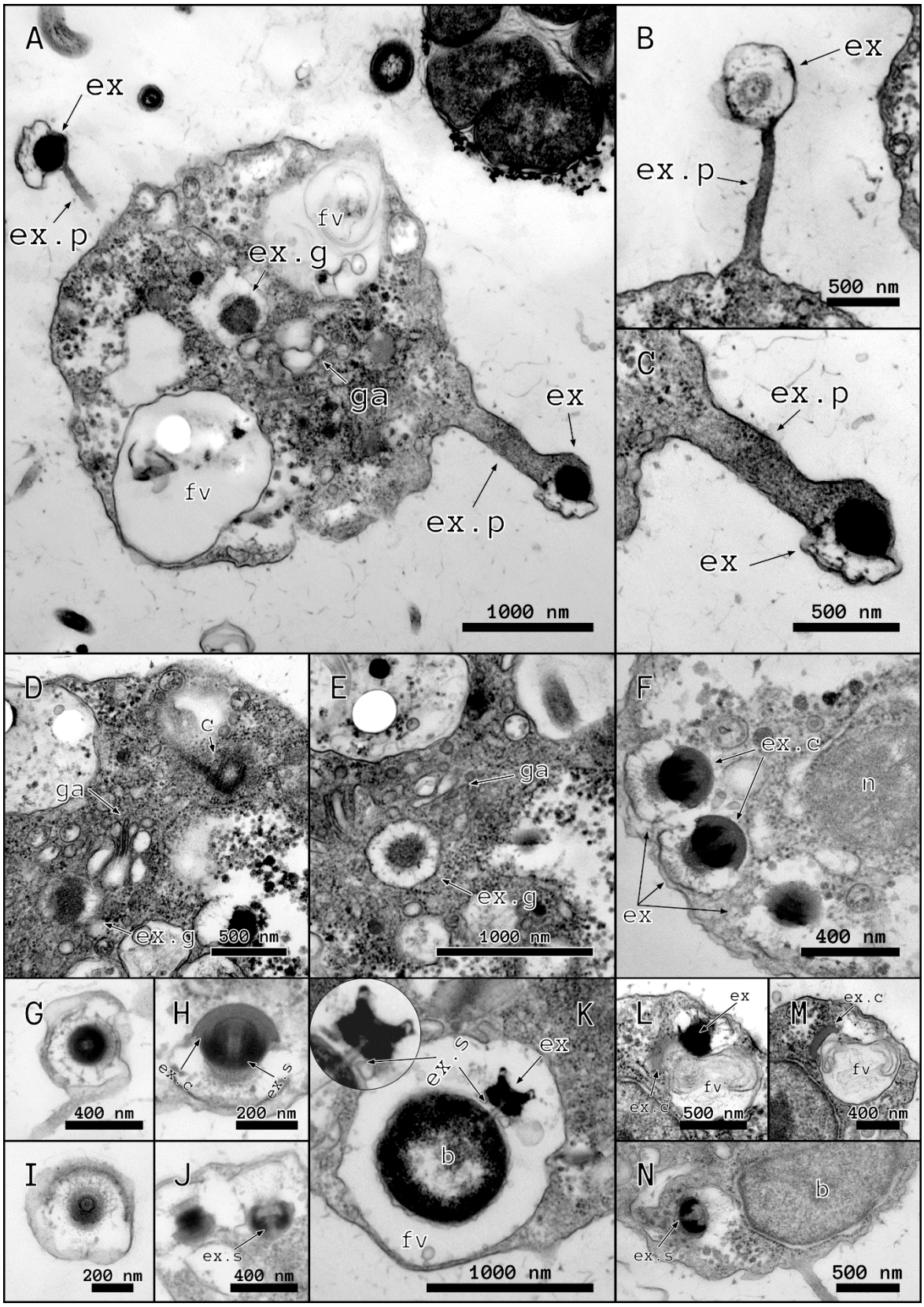

Figure 4: TEM micrographs of extrusomes of sun-like stages of organism SUM-K. **A**: representative cell with two discharged extrusome projections and putative extrusome genesis near Golgi apparatus. **B, C**: details on extrusome projections and discharged extrusome body; longitudinal (B) and cross-section (C). **D, E**: serial sections of the same

cell showing genesis of the extrusome body in Golgi apparatus. **F**: typical arrangement of cross-sectioned extrusive bodies inside the cell, under surface of the cell membrane. **G, I**: longitudinal cuts of discharged extrusomes with high magnification. **H, J**: high magnification of cross-sectioned extrusome bodies with visible extrusome shaft and cap. **K**: visible connections between extrusome body and prokaryotic prey, in circle is enlarged detailed display. **L, M**: disconnected extrusome cap and body depicted by consecutive serial sections of the same cell. **N**: detail on distinctive morphology of extrusome body. Internal morphology of extrusome body is best shown in microtomes F, H and I. (b – bacterial prey; c – centriole; ex – extrusome body; ex.c – extrusome cap; ex.g – extrusome genesis; ex.p – extrusome projection; ex.s – extrusome shaft; fv – food vacuole; ga – Golgi apparatus; n – nucleus).

TEM microsections they appear to have the same shape as "disconnected" extrusome caps of SUM-K (Mylnikov and Mylnikov, 2014). In the case of the kinetocysts of the cerozoan *Thaumatomastix* sp. these extrusomes are clearly supported by microtubules, are more oval and have numerous shafts spanning through the extrusome body (Zolotarev et al., 2011). Similarly, the telonemosomes of *Telonema subtilis* are more oval and have wheel-like structure on cross-section that is made up by 13 components, each of which has a hole in it (Yabuki et al., 2013). This is the same number of filaments as in extrusomes of cerozoan *Metromonas simplex*, described as trichocysts. These organelles show clear compartmentalization into 13 sections bounded by each filament (Myl'nikova and Myl'nikov, 2011). Lastly, *Jakoba libera* has extrusomes of similar, but more oval, shape that might be also hollow, but their description is only brief (Patterson, 1990).

Heliozoans are known for their radiating, microtubule-supported axopodia and have shapes reminiscent of the sun-like stage of SUM-K. The greatest resemblance can be observed at a distinctive axopodium of *Microheliella maris*. This putative heliozoan (molecular data are still missing) have extrusomes that are produced at the tip and along the length of the axopodia but other than the same circular shape they lack resemblance (Yabuki et al., 2012). Other heliozoan extrusomes, kinetocysts, were described in *Raphidiophrys contractilis* (Sakaguchi et al., 2002). These extrusomes have the same internal compartmentalization as proposed for SUM-K extrusomes i.e., extrusome cap, hollow channel, and electron dense core. Lastly, a relative of SUM-K (see next chapter), *Ancoracysta twista*, has only one big extrusome (ancoracyst) per cell, that is ultrastructurally distinct from the extrusomes of SUM-K, particularly its ovate shape, internal symmetry of compartments, and ultrastructure of the anchor-shaped extrusome cap (Janouškovec et al., 2017).

3.3. Phylogenetic position of SUM-K

The SSU rRNA gene is a commonly used marker for determination of phylogenetic relationships of eukaryotes (e.g., Marande et al., 2009; Holzmann and Pawlowski, 2017; Arroyo et al., 2018). This molecular marker is still the gold standard for eukaryotic phylogeny, even 50 years after its introduction (Fox et al., 1977). In our case, this single gene analysis failed for the most part and was unable to resolve the phylogenetic position of SUM-K. This, in itself, is a strong indication that this organism may form a deep-branching eukaryotic lineage (Pawlowski et al., 2012). The position of SUM-K on the phylogenetic tree varied depending on the models used and the sampling of eukaryotic lineages but was never sufficiently statistically supported. Moreover, SUM-K formed one of the shortest branches which also suggests its peculiar phylogenomic position (Fig. 5) (complete tree available at Suppl. fig. 5). Even though statistical support was very low, this analysis placed SUM-K as a relative of *A. twista*. This position was later confirmed by phylogenomic analysis showing that the SSU rRNA gene is rightfully regarded as the best single-gene molecular marker for establishment of evolutionary relationships.

A dataset containing orthologous proteins of representatives of the main groups of eukaryotes has been used many times for phylogenomic inference of newly discovered eukaryotes (e.g., Kang et al., 2017; Lax et al., 2018; Lahr et al., 2019). Phylogenomic analysis of SUM-K showed a close evolutionary relationship of this organism with supergroup Hemimastigophora and orphan lineage *Ancoracysta* with full statistical support (Fig. 6).

The first Hemimastigophorean, *Spironema multiciliatum*, was recognized by its unusual morphology already in 1892 (Klebs). Forgotten for a century, it acquired phylum affiliation of *incertae sedis* only after description of the ultrastructure of *Hemimastix amphikineta* (Foissner et al., 1988). But, for the first molecular data, we had to wait another two decades until *Hemimastix kukwesjijk* and *Spironema cf. multiciliatum* were comprehensively described, promoting the phylum to the, currently, newest eukaryotic supergroup of uncertain phylogenomic position, Hemimastigophora (Lax et al., 2018). Orphan lineage Ancoracysta, defined by a single species, *Ancoracysta twista*, was also discovered only recently, being placed as a basal lineage of Haptophyta (Janouškovec et al., 2017). Notably, both of these publications lacked data from each other but even comprehensive phylogenomic studies with broad sampling (including both taxons) that were

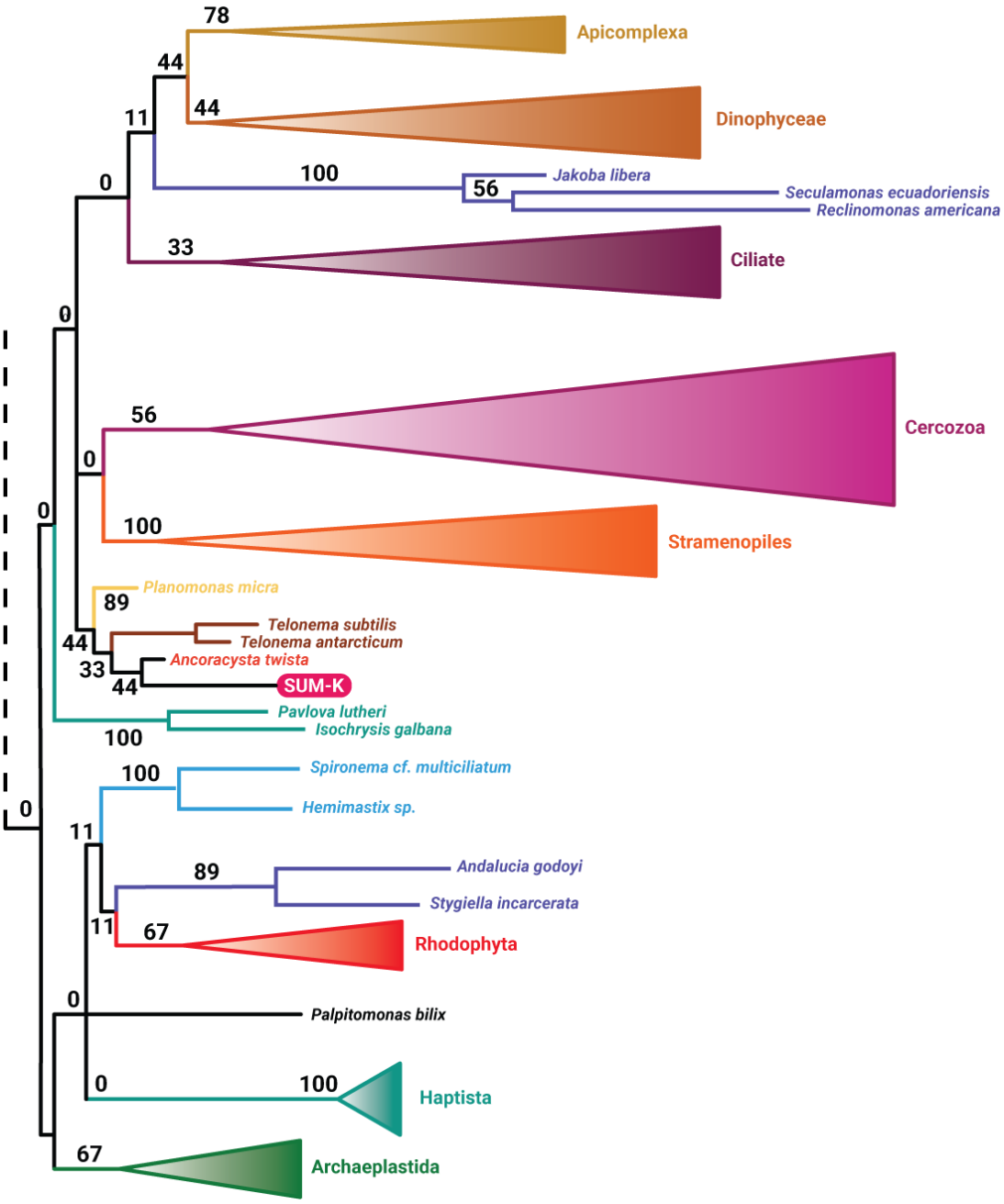

Figure 5: Phylogenetic tree of eukaryotes based on gene sequences for SSU rRNA, created by the Maximum likelihood method in the program RAxML with model GTRGAMMAI (Stamatakis, 2006) and statistical topology support estimation by non-parametric bootstraps with 100 replicates, rooted by midpoint method (root not shown). For clarity, only a small segment comprised of clade with SUM-K is shown while a full, edited phylogenetic tree is available at Suppl. fig. 5. The individual eukaryotic lineages are merged if they contain more than 3 representatives. Statistical support for most lineages on the tree is very low and is reported as bootstrap value on the branches. On this tree, SUM-K is placed close to the *Ancoracysta* lineage, but this position is not statistically supported. Modified in Adobe Illustrator.

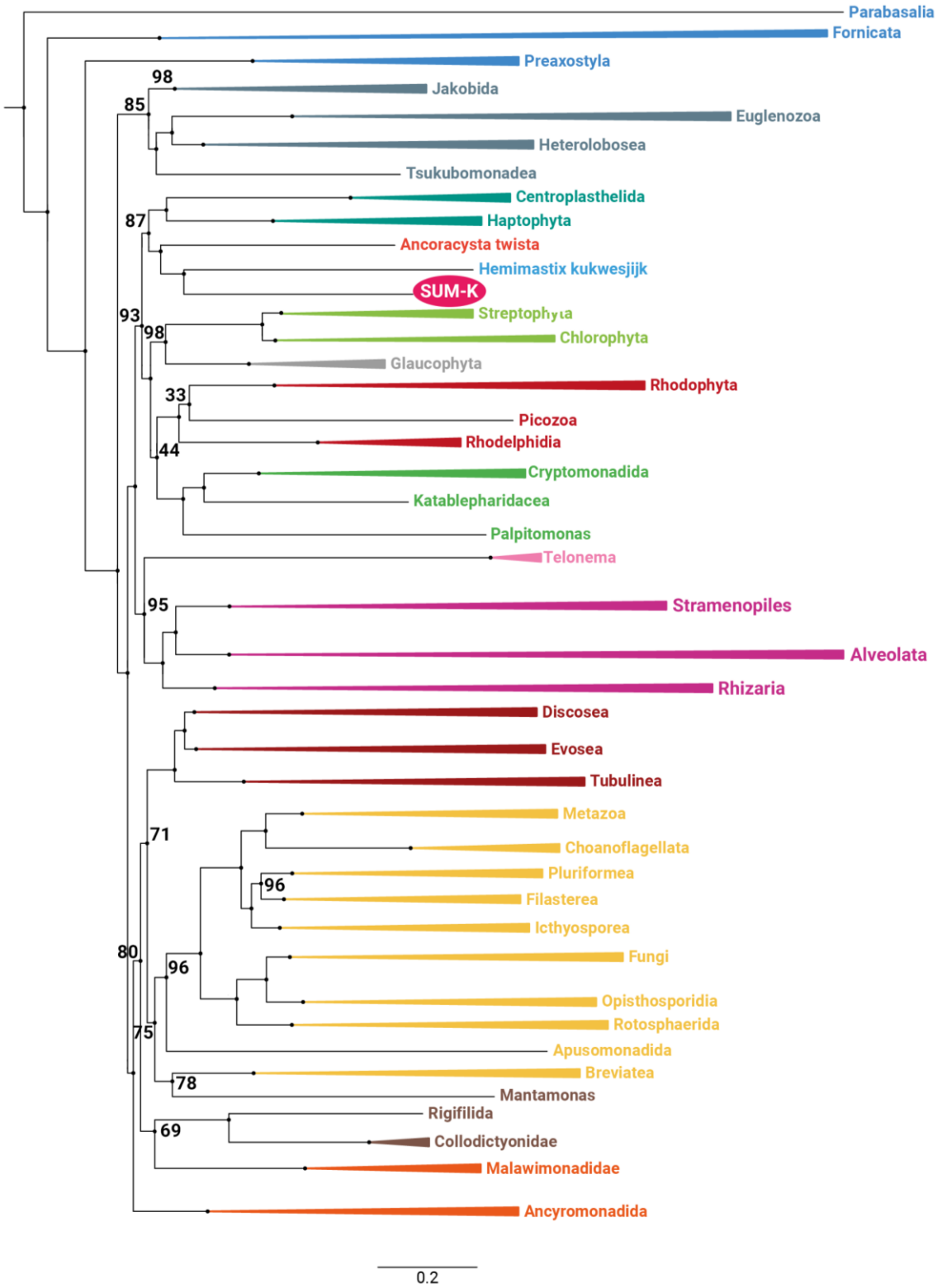

Figure 6: Inference of evolutionary relationships of organism SUM-K by a phylogenomic tree of 252 eukaryotes, inferred from 206 orthologs using ML method with LG+Γ4+C60-PMSF model, and an LG+ Γ4+C20 ML tree as a PMSF guide input tree in software IQ-

TREE v2.0-rc1 (Minh et al., 2020). MLBS were inferred from collected PMSFs from the above-described ML analysis and used for non-parametric bootstraps with 100 replicates. Final supermatrix of 61995 AA sites as well as single-protein alignments and subsequent supermatrix-based phylogenomic analysis were processed with the PhyloFisher software with settings as recommended by Tice et al. (2021). MLBS values of 100% are not shown, all other values are indicated at their respective node. Root is placed on parabasalid *Trichomonas vaginalis* by midpoint method. For the sake of clarity, individual eukaryotic lineages are merged and colored by their supergroup affiliation. Length of the branches represents average number of substitutions per site and can be determined by scale included. The most recent nomenclature was used and appropriated from Adl et al. (2019). Graphical edit was made in Adobe Illustrator.

aimed at resolving evolutionary relationships of eukaryotic lineages failed to define the place of Ancoracysta and Hemimastigophora (Burki et al., 2020; Schön et al., 2021). Our phylogenomic analysis managed just that, on the top of it with full statistical support.

Other notable outcomes of this phylogenomic analysis are the positions of Picozoa and Mantamonas lineages. The former lineage was initially discovered in marine environmental clone libraries and its evolutionary position was established only very recently (Schön et al., 2021). Together with Rhodelphidia, a sister lineage to red algae (Gawryluk et al., 2019), they changed our view on early evolution of supergroup Archaeplastida as both of these lineages lack a photosynthetic chloroplast, rejecting the widely-assumed theory that Archaeplastids evolved from a photosynthetic common ancestor and, instead, implying that the first Archaeplastid was mixotrophic. Furthermore, monophyly of the whole supergroup Archaeplastida is coming into question, with some phylogenomic analyses placing Cryptista inside of it as a as sister lineage to Rhodophyta (Lax et al., 2018; Gawryluk et al., 2019; Strassert et al., 2019) and others immediately rejecting this topology (Irisarri et al., 2021; Strassert et al., 2021). Our phylogenomic analysis suggested the first alternative but, statistical support is very low.

Lineage Mantamonas is, at the current state of knowledge, a defining member of supergroup CRuMs (Brown et al., 2018). Our phylogenomic analysis did not support this and, instead, affiliated lineage Mantamonas as a basal group of Obazoa. This topology is rather intriguing as Mantamonas, CRuMs and whole Obazoa are placed on opposite sides of eToL from SUM-K, where relations of supergroups should presumably not be altered in a significant way. Nevertheless, CRuMs cannot survive without M and CRus just does not have the same ring, but before we send this supergroup to the cemetery of former

nomenclature, we will need additional proof that what we see is, indeed, true topology and not a result of some unappreciated artefact.

The graphically edited phylogenomic tree with visualized supergroup affiliations and showing all organisms that were used in this analysis is available at Suppl. fig. 6. Also, see Suppl. fig. 7 for illustrative depiction of eToL inferred from this analysis.

3.4. Mitochondria and mitogenome

The key elements predicating diversity of mitochondrial metabolism are lineage-specific innovations, retentions, and new acquisitions of traits in connection with a varying degree of gene losses from its proto-mitochondrial ancestor (Pyrih et al., 2021). (Pyrih et al., 2021). Therefore, characterizing the genes directly encoded by mtDNA is essential. What follows is the characterization of a complete mitogenome of organism SUM-K (Fig. 7). With the size of 43 872 bp and 64 protein coding sequences it belongs to average sized mitogenomes. Noteworthy is a presence of SSU mitoribosomal proteins (*rps2, rps7, rps10, rps11, rps12, rps13, rps14*, and *rps19*) and LSU mitoribosomal proteins (*rpl6, rpl11, rpl14*, and *rpl16*). Even though they are present in various protists and plants, they belong to the group of extended mitochondrial gene sets as described by Grey at al. (2004). Similarly, subunits of the ETC *sdh2* (complex II), *atp1*, and *atp3* (complex V) are found only in protist lineages and were transferred into the nuclear genome in animals, plants, and fungi. Interestingly, trypanosomatids and euglenids transferred only part of the *sdh2* gene and a small fragment remained encoded in the mitochondrion (Morales et al., 2009). Also remarkable is the presence of *cox11* – a subunit of complex IV, known only from *Naegleria gruberi* and jakobids (Gray et al., 2004; Burger et al., 2013). Even though all of these proteins were previously demonstrated to be present in protist mitochondria it still shows substantial complexity of SUM-K mitochondrial metabolic activity. All besides one.

Here, I present the first recorded occurrence of mitochondrially-encoded secA protein, demonstrating retention of the plesiomorphic alphaproteobacterial metabolic pathway that was lost in all other eukaryotes (Fig. 8). As expected, all photosynthetic eukaryotes formed solid clade. The same applies for non-photosynthetic protists with reduced chloroplast where secA logically diverged more, as it lost its original function of transport of photosynthesis metabolites. Separate clade was also formed for duplicated gene secA in land plants. SUM-Ks mitochondrially encoded secA formed one of the longest branches near the base of the tree. Each sequence of eukaryotes was manually checked for

domain architecture in InterPro scan (Jones et al., 2014) to avoid possible missannotations in public databases.

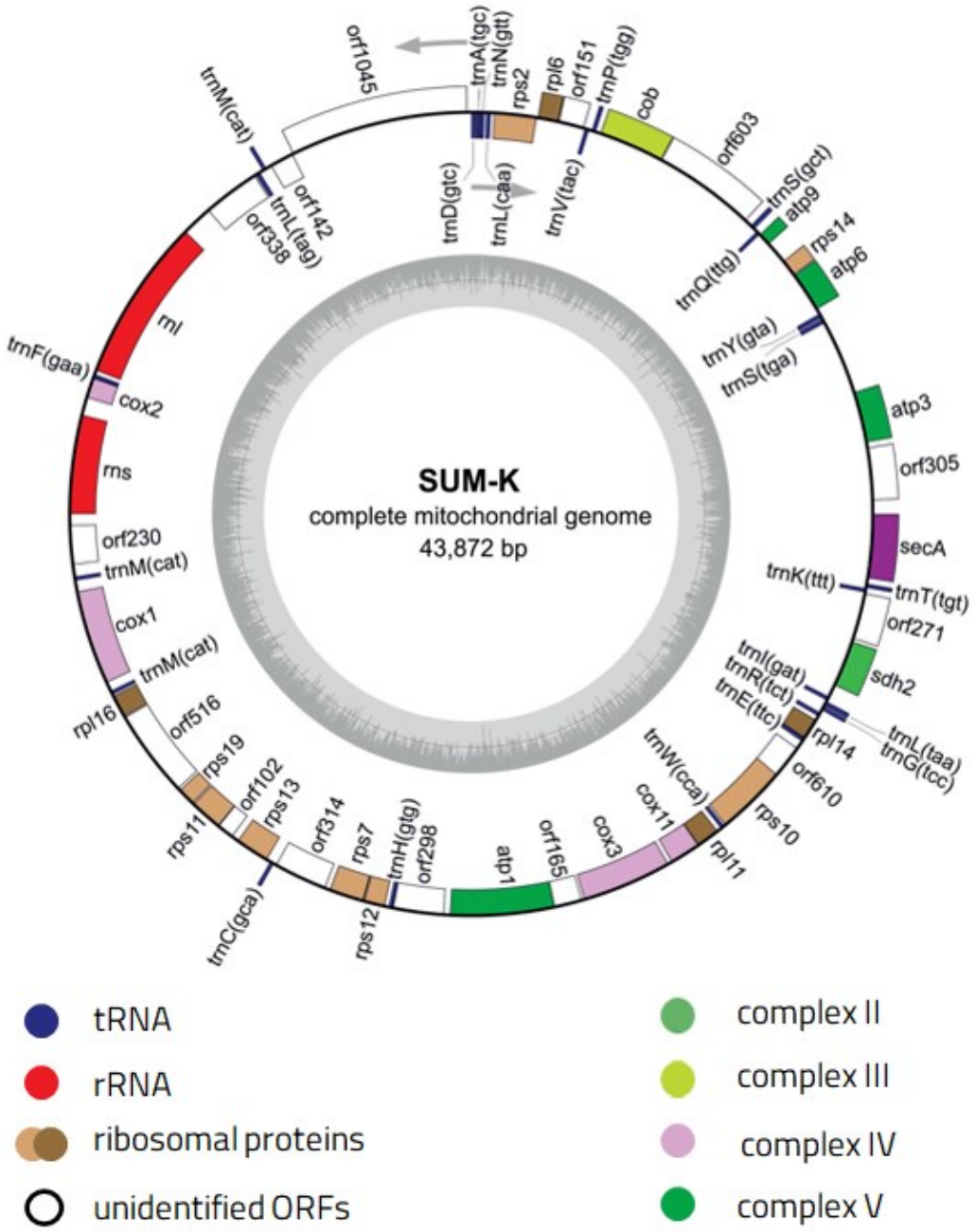

Figure 7: Predicted mitochondrial genome of organism SUM-K. Prediction was performed with utility MFannot (Lang and Burger; (unpublished)) and graphical representation was made in OrganellarGenomeDRAW toolkit (Greiner et al., 2019), later modified in Adobe Photoshop. Most noteworthy is the peculiar presence of a set of genes encoding subunits of ETC complexes and the first-ever observed mitochondrially encoded gene secA (purple).

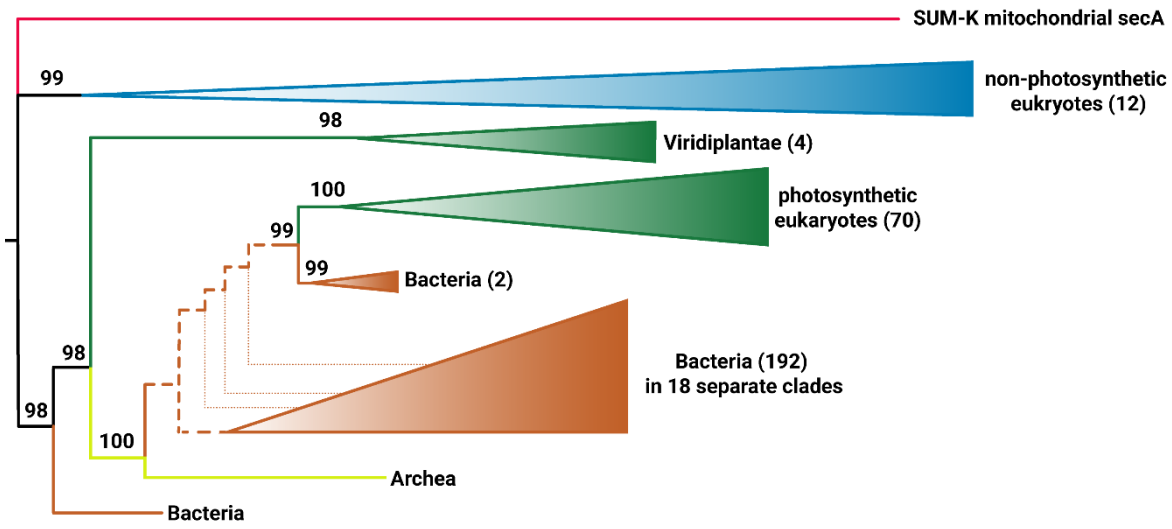

Figure 8: Unrooted phylogenetic tree of gene secA, inferred from bacterial, archaeal and eukaryotic sequences. Tree was created with Maximum Likelihood method in the program IQ-TREE (version 1.6.12; Nguyen et al., 2015) with 10000 ultrafast bootstrap approximations (Hoang et al., 2018) that are reported on branches. Model used was LG4X, protein mixture models with FreeRate heterogeneity (Le et al., 2012). Dataset contains two best eukaryotic BLAST hits of SUM-K secA gene from Tara environmental metagenomic project (Sunagawa et al., 2015), that branched within clade of photosynthetic eukaryotes. For clarity, only relevant and merged clades are shown. In parenthesis are for each clade reported counts of taxa.

In bacteria, there are two main protein secretory pathways: the TAT (twin-arginine translocation) system was described in diverse eukaryotic lineages and is often taken as an example of gradual loss of alphaproteobacterial pathways in mitochondria (Petru et al., 2018); the second is the Sec-dependent protein secretion pathway with the SecYEG translocon complex that is mediated by essential ATPase secA (Cranford-Smith and Huber, 2018). In eukaryotes, remnants of this pathway remained only on a periphery. Mostly, they are present in the photosynthetic organisms as an acquisition from chloroplast endosymbiosis. In glaucophytes, secA is exclusively encoded in the nuclear genome, such as in *Cyanophora paradoxa* (Steiner et al., 2012), or is absent altogether (Kim et al., 2020). In cryptophytes, on the other hand, it can be found only in the plastid genome (Kim et al., 2017) and has been retained even in the non-photosynthetic species *Cryptomonas paramecium* (Hadariová et al., 2018). In rhodophytes the issue is more complex. A majority of red algae have secA encoded only in the plastid genomes (Figueroa-Martinez et al., 2019), but there are few exceptions, mostly plesiomorphic extremophiles, that have a secA gene homolog duplicated in both the chloroplast and nuclear genomes (e.g., *Cyanidioschyzon*

merolae; Koyama et al., 2011). The occurrence in Stramenopiles and Haptista is comparable to rhodophytes but with several cases of deletion, consistent with the hypothesis of ongoing plastid genome reduction in these lineages (Kim et al., 2020). Finally, green algae and land plants took gene duplication and EGT to the next level and have both copies of secA encoded in their nucleus (Fernandez, 2018).

The second occurrence of a Sec-dependent protein secretion pathway in eukaryotes is in endoplasmic reticulum-associated protein transport, but this pathway does not involve secA as ATPase so it will not be discussed further (Sun et al., 2007).

It needs to be noted that secA was not found in our transcriptomic data. This is possibly due to a known artefact of mRNA sequencing with the poly-A selection method of transcript enrichment, where mRNAs that are non-polyadenylated (absent of a poly-A tail at their 3' ends), are not captured in the final library (Yang et al., 2011; Morlan et al., 2012).

Polyadenylation of mitochondrial encoded mt-mRNAs is complex issue. It would be logical to assume that similarly to their alphaproteobacterial ancestor polyadenylation in mitochondria leads to rapid decay of transcripts by nucleases and associated factors (Hajnsdorf and Kaberdin, 2018). But in eukaryotes it is not the case. Different evolutionary lineages approach polyadenylation of mt-mRNAs differently and often have contrasting role in regards to stability, e.g., human mt-mRNAs are polyadenylated exclusively (Temperley et al., 2003), in yeasts it is completely dispensable in regards of protein maturation, and land plants use it for facilitating mRNA decay (Gagliardi et al., 2004). From protist world, *Trypanosoma brucei* can either use polyadenylation to mark proteins for decay or, depending on the length of poly-A tail, stabilization of the mRNAs transcripts (Aphasizheva and Aphasizhev, 2021). To date, nothing is known about polyadenylation of mt-mRNAs in Hemimastigophoreans but based on the lack of mitochondrially encoded proteins in our transcriptome data and based on our hypothesis that SUM-K has rather plesiomorphic mitochondria, we would assume that polyadenylation of mt-mRNAs does not occur for purpose of protein maturation, similarly to its alphaproteobacterial ancestors. Hence, the effect - absent mt-mRNA transcripts in our data.

At this point predicting function of mitochondrially encoded secA is speculative. Based on near universal conservation of secA in bacteria (Cranford-Smith and Huber, 2018) and retention of similar functions in chloroplast and photosynthetic organisms it is safe to assume, that even mitochondrially encoded secA preserved its function of the protein secretion over the cytoplasmatic membrane – or this case mitochondrial membrane.

Moreover, presence of its usual partners, mainly secB, secY and the rest of the secYEG translocon most of which were observed in various protist lineages (e.g., in jakobids mitochondrial genome (Gray et al., 2004)) is disputable. EggNOG automatic annotator was able to identify secYEG homolog, but it appears to be rather divergent and more comprehensive phylogenetic analysis is required. Furthermore, even if we found protein transport channel that secA binds to, it does not solve main question - why was this secretory system entirely abandoned in rest of the eukaryotes and why it is retained in SUM-K. Further steps should also involve prediction of internally encoded targeting signal recognized by Sec machinery, (as it is in bacteria (Cranford-Smith & Huber, 2018)) and subsequently proteins that carry it. This could potentially cast some light into rationale of this system.

3.5 Predictions of ETC subunits and energy metabolism enzymes

As SUM-K is routinely cultured in an anoxic environment it can be presumed that it is fully adapted to life without oxygen. At the same time, its mitochondrial genome is unusually complex with retentions of some rare subunits of ETC hinting that this organism has an extremely versatile mitochondrial metabolism. For detailed characterization we annotated all subunits of ETC and associated assembly proteins (Tab. 1).

At the first glace it seems that SUM-K bears all complexes of ETC with at least several identified subunits. Only by the close examination, we can conclude that few complexes show typical reduction as shown by the variety of anaerobic protists (Müller et al., 2012; Stairs et al., 2015; Gawryluk and Stairs, 2021). Complex I (CI - NADH:ubiquinone oxidoreductase) is the most obvious example where only two core subunits were identified, NuoF and NuoE. Complementary subunit NDUFAB1 is a multifunctional protein with activity in several different enzymes (Valach et al., 2018) and assembly factor NDUFAF6, that have unclear function, is putatively also connected to various metabolic pathways (Cogliati et al., 2018). NuoF and NuoE are, outside of being core subunits of CI, also speculated to be involved in hydrogen production of hydrogenosomes of anaerobic protists (Schut and Adams, 2009; Gawryluk and Stairs, 2021). In this pathway, they are supposedly connected to [FeFe] hydrogenase (HydA) where they shift reaction balance in favor of ferredoxin reduction, making HydA more efficient. For the proper assembly, HydA requires tree maturases HydE, HydF and HydG while initial oxidation of ferredoxin is mediated by oxygen–sensitive pyruvate:ferredoxin oxidoreductase (PFO) (Gawryluk and Stairs, 2021). All of these enzymes were manually annotated in the SUM-K transcriptome (Tab. 2),

suggesting that CI is completely absent in SUM-K, and the annotated CI subunits are parts of different energy metabolism pathways.

Second complex (CII - Succinate dehydrogenase, SDH) is usually composed of 4 distinctive subunits, each with its respective assembly factor. The core subunits, SDH1 and SDH2, were observed to form soluble assembly detached from mitochondrial inner membrane, where rest of the subunits are anchored (Bezawork-Geleta et al., 2017; Moosavi et al., 2019). Same situation is proposed for mitochondrion of SUM-K as we were able to annotate two main subunits and assembly factor 1. The assembly factor 2 (flavinator of SDH) was also observed, but the sequence was fragmented and therefore is not reported. Notably, we also identified enzyme fumarate reductase (FRD) which is important component of anaerobic respiration that catalyzes the reverse reactions of SDH (Jardim-Messeder et al., 2017). The rest of the ETC of SUM-K (CIII - cytochrome c oxidoreductase, CIV - cytochrome c oxidase, CV - F1Fo ATP synthase) appears to be complete or near complete. CIII is lacking a few subunits of low molecular size and several assembly proteins, but has all three main components present (Cyc1, cob, ISP). On the other hand, CIV contains a rich set of assembly factors and associated proteins together with majority of core subunits. CV is seemingly also near complete, which is surprising in the case of a putative anaerobe. It has been proposed that CV in anaerobic protists is only a remnant of complex metabolism of the ancestral mitochondria that acts as rudimentary ATP source, and is destined to degrade in evolutionary process (Gawryluk et al., 2016; Lewis et al., 2020).

Last aspect I cannot omit while describing the energy metabolism of SUM-K is presence of alternative NAD(P)H dehydrogenase (NDA) and alternative oxidase (AOX), which provide auxiliary pathway of membrane potential generation to CI (Antos-Krzeminska and Jarmuszkiewicz, 2019; Juergens et al., 2021), reportedly with comparable efficiency (Antos-Krzeminska and Jarmuszkiewicz, 2014).

Table 1: Predictions of subunits of ETC complexes of organism SUM-K. The first column contains name of a subunit, second shows used shortcuts, and in third is method of annotation by which prediction was made. For clarity, every complex is separated.

Complex I: NADH:ubiquinone oxidoreductase		
subunit 51-kDa	NDUFV1/NuoF	KEGG
subunit 24-kDa	NDUFV2/NuoE	KEGG
subunit 8kDa acyl-carrier protein	NDUFAB1	KEGG
assembly factor 6	NDUFAF6	EggNOG

Complex II: Succinate dehydrogenase		
flavoprotein subunit	SDH1	KEGG
iron-sulfur subunit	SDH2	MFannot
assembly factor 1	SdhAF1	manual
Complex III: cytochrome c oxidoreductase		
heme protein	Cyc1	KEGG
apocytochrome b	cob	MFannot
subunit Rieske iron-sulphur protein	ISP	KEGG
subunit 6, 7.8-kDa hinge protein	QCR6	manual
subunit 7, 14-kDa subunit	QCR7	KEGG
assembly factor 1	UQCC1/Cbp3	manual
Complex IV: cytochrome c oxidase		
subunit 1	COX1	MFannot
subunit 2	COX2	MFannot
subunit 3	COX3	MFannot
subunit 5b	COX5B	KEGG
subunit 6a	COX6A	manual
subunit 6b	COX6B	manual
assembly protein, Protohaem IX farnesyltransferase	COX10	manual
assembly protein	COX11	MFannot
assembly factor, HemeA synthase	COX15	manual
assembly factor, copper chaperone	COX17	manual
assembly factor, mitochondrial inner membrane protein	COX18/Oxa1L	manual
assembly factor	COX19	manual
assembly factor	COX20	manual
assembly factor, mt-mRNA-processing protein	COX24	manual
assembly factor 4	Coa4	manual
Complex V: F1Fo ATP synthase		
F1 subunit α	Atp1/AtpA	MFannot
F1 subunit β	Atp2/AtpD	KEGG
F1 subunit γ	Atp3/AtpG	MFannot
F1 subunit δ	Atp16/AtpC	KEGG
F1 subunit ε	Atp15	KEGG

F1 complex assembly factor 2	ATPAF2/Atp12	EggNOG
Fo subunit OSCP	Atp5/AtpH	KEGG
Fo subunit a	Atp6/AtpB	MFannot
Fo subunit c	Atp9/AtpE	MFannot

Table 2: List of selected enzymes involved in energy metabolism of organism SUM-K. Every enzyme in the list was annotated manually and analyzed by single gene phylogenetic analysis. For KEGG annotated and curated proteins with prediction of mitochondrial targeting sequences see Supplementary table 3.

ACH	acetyl-CoA hydrolase
ACS ab1	acyl-CoA synthetase
ACS ab2	acyl-CoA synthetase
AOX	alternative oxidase
ASCT 1A	acetate:succinate CoA transferase 1A
ASCT 1B	acetate:succinate CoA transferase 1B
ASCT 1C	acetate:succinate CoA transferase 1C
CytC_1A-B	cytochrome c, class IA/ IB
cytGAPDH	glyceraldehyde 3-phosphate dehydrogenase (cytoplasmatic)
ENO	phosphopyruvate hydratase (enolase)
FBA I	fructose-1,6-bisphosphate aldolase class I
Fd	ferredoxin
FRD	fumarate reductase
GPI	glucose-6-phosphate isomerase
HEX-FRK	fructokinase
HEX-GLK	glucokinase
HydA	[FeFe] hydrogenase
HydE	[FeFe] hydrogenase H-cluster radical SAM maturase
HydF	[FeFe] hydrogenase H-cluster maturation GTPase HydF
HydG	[FeFe] hydrogenase H-cluster radical SAM maturase HydG
iPGM	cofactor-independent phosphoglycerate mutase
LSC	succinate—CoA ligase
Me	malic enzyme
NDA1	alternative NAD(P)H-ubiquinone oxidoreductase A1
PFK	phosphofructokinase
PFL	pyruvate formate lyase
PFLA	pyruvate formate-lyase activating enzyme
PFO	pyruvate:ferredoxin oxidoreductase
PFP	phosphotransferase
PGK	phosphoglycerate kinase
TPI	triosephosphate isomerase

4. Conclusions

The SUM-K (or as we familiarly call it kapička-kulička-slzička) is indisputably fascinating. From its peculiar morphology, through extraordinary ultrastructure with novel described organelle, to unique retention of ancestral pathway in mitogenome and unexpectedly versatile mitochondrial metabolism, it SUM-K keeps surprising us.

Because we were not able to find any TEM sections of the flagellated stage of SUM-K, we believe the next one is waiting right there. We will not only search for flagellar apparatus and associated cytoskeletal elements, but also the structure of the intriguing posterior extrusion. It could be a modified flagellum, pseudopodium, or, perhaps, some undescribed structure with a peculiar function. For this, we are planning to experiment with differential and gradient centrifugation. If we manage to separate flagellated stages from the rest of the cells, it can be pivotal milestone from which we can start thinking about employing comparative transcriptomics, that could lay open the meaning of this dualistic morphology.

Phylogenomic analysis established evolutionary relationship of SUM-K and *A. twista*, placing them inside the supergroup Hemimastigophora, with full statistical support. It remains to be seen if the group just got bigger by two new fellows or another name will be proposed. Some progress will almost certainly come after long awaited publication of phylogenomic analysis of *Meteora sporadica* and "protist X", that are seemingly next in line for admission to this supergroup (presented by Yana Eglit, 1st Electronic Symposium on Protistology, 2021). Hopefully, by including these two bizarre protists into the phylogenomic dataset we will also resolve relations of Mantamonas and Cryptista, which placement on our eToL is almost certainly result of computational artefact.

Characterization of energy metabolism and ETC composition of SUM-K was just first step in annotation of transcriptome. It showed that SUM-K uses the vestiges of its aerobic machinery to tolerate transient exposure to oxygen and described a rich set of enzymes of anaerobic metabolism. Next, we will aim at other metabolic pathways, certainly at the Sec-dependent secretion that needs to be examined to the details, but also other enzymes of great evolutionary importance, like the myosin cytoskeletal motors or the adaptor protein complexes.

Finally, we will soon be standing before the most challenging task of them all. Finding a proper name.

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., CÁRDENAS, P., ČEPIČKA, I., CHISTYAKOVA, L., DEL CAMPO, J., DUNTHORN, M., EDVARDSEN, B., EGLIT, Y., GUILLOU, L., HAMPL, V., HEISS, A. A., HOPPENRATH, M., JAMES, T. Y., KARNKOWSKA, A., KARPOV, S., KIM, E., KOLISKO, M., KUDRYAVTSEV, A., LAHR, D. J. G., LARA, E., LE GALL, L., LYNN, D. H., MANN, D. G., MASSANA, R., MITCHELL, E. A. D., MORROW, C., PARK, J. S., PAWLOWSKI, J. W., POWELL, M. J., RICHTER, D. J., RUECKERT, S., SHADWICK, L., SHIMANO, S., SPIEGEL, F. W., TORRUELLA, G., YOUSSEF, N., ZLATOGURSKY, V., ZHANG, Q. 2019. Revisions to the Classification, Nomenclature, and Diversity of Eukaryotes. In: *Journal of Eukaryotic Microbiology*. 2019, vol. 66, no. 1, pp. 4–119. ISSN 15507408. Available at: DOI: 10.1111/jeu.12691

AGARWALA, R., BARRETT, T., BECK, J., BENSON, D. A., BOLLIN, C., BOLTON, E., BOUREXIS, D., BRISTER, J. R., BRYANT, S. H., CANESE, K., CHAROWHAS, C., CLARK, K., DICUCCIO, M., DONDOSHANSKY, I., FEDERHEN, S., FEOLO, M., FUNK, K., GEER, L. Y., GORELENKOV, V., HOEPPNER, M., HOLMES, B., JOHNSON, M., KHOTOMLIANSKI, V., KIMCHI, A., KIMELMAN, M., KITTS, P., KLIMKE, W., KRASNOV, S., KUZNETSOV, A., LANDRUM, M. J., LANDSMAN, D., LEE, J. M., LIPMAN, D. J., LU, Z., MADDEN, T. L., MADEJ, T., MARCHLER-BAUER, A., KARSCH-MIZRACHI, I., MURPHY, T., ORRIS, R., OSTELL, J., O'SULLIVAN, C., PANCHENKO, A., PHAN, L., PREUSS, D., PRUITT, K. D., RODARMER, K., RUBINSTEIN, W., SAYERS, E., SCHNEIDER, V., SCHULER, G. D., SHERRY, S. T., SIROTKIN, K., SIYAN, K., SLOTTA, D., SOBOLEVA, A., SOUSSOV, V., STARCHENKO, G., TATUSOVA, T. A., TODOROV, K., TRAWICK, B. W., VAKATOV, D., WANG, Y., WARD, M., WILBUR, W. J., YASCHENKO, E., ZBICZ, K. 2016. Database resources of the National Center for Biotechnology Information. In: *Nucleic Acids Research*. 2016, vol. 44, no. D1, pp. D7–D19. ISSN 13624962. Available at: DOI: 10.1093/nar/gkv1290

ANTOS-KRZEMINSKA, N., JARMUSZKIEWICZ, W. 2014. Functional expression of the Acanthamoeba castellanii alternative oxidase in Escherichia coli; Regulation of the activity and evidence for Acaox gene function. In: *Biochemistry and Cell Biology*. 2014, vol. 92, no. 3, pp. 235–241. ISSN 12086002. Available at: DOI: 10.1139/bcb-2014-0014

ANTOS-KRZEMINSKA, N., JARMUSZKIEWICZ, W. 2019. Alternative Type II NAD(P)H Dehydrogenases in the Mitochondria of Protists and Fungi. In: *Protist*. 2019, vol. 170, no. 1, pp. 21–37. ISSN 16180941. Available at: DOI: 10.1016/j.protis.2018.11.001

APHASIZHEVA, I., APHASIZHEV, R. 2021. Mitochondrial RNA quality control in trypanosomes. In: *Wiley Interdisciplinary Reviews: RNA*. 2021, vol. 12, no. 3, pp. 1–16. ISSN 17577012. Available at: DOI: 10.1002/wrna.1638

ARAMAKI, T., BLANC-MATHIEU, R., ENDO, H., OHKUBO, K., KANEHISA, M., GOTO, S., OGATA, H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. In: *Bioinformatics*. 2020, vol. 36, no. 7, pp. 2251–2252. ISSN 14602059. Available at: DOI: 10.1093/bioinformatics/btz859

ARMENTEROS, J. J. A., SALVATORE, M., EMANUELSSON, O., WINTHER, O., VON HEIJNE, G., ELOFSSON, A., NIELSEN, H. 2019. Detecting sequence signals in targeting peptides using deep learning. In: *Life Science Alliance*. 2019, vol. 2, no. 5, pp. 1–14. ISSN 25751077. Available at: DOI: 10.26508/lsa.201900429

ARROYO, A. S., LÓPEZ-ESCARDÓ, D., KIM, E., RUIZ-TRILLO, I., NAJLE, S. R. 2018. Novel diversity of deeply branching holomycota and unicellular holozoans revealed by metabarcoding in Middle Paraná River, Argentina. In: *Frontiers in Ecology and Evolution*. 2018, vol. 6, no. JUL, pp. 1–17. ISSN 2296701X. Available at: DOI: 10.3389/fevo.2018.00099

BATEMAN, A., MARTIN, M. J., ORCHARD, S., MAGRANE, M., AGIVETOVA, R., AHMAD, S., ALPI, E., BOWLER-BARNETT, E. H., BRITTO, R., BURSTEINAS, B., BYE-A-JEE, H., COETZEE, R., CUKURA, A., SILVA, A. Da, DENNY, P., DOGAN, T., EBENEZER, T. G., FAN, J., CASTRO, L. G., GARMIRI, P., GEORGHIOU, G., GONZALES, L., HATTON-ELLIS, E., HUSSEIN, A., IGNATCHENKO, A., INSANA, G., ISHTIAQ, R., JOKINEN, P., JOSHI, V., JYOTHI, D., LOCK, A., LOPEZ, R., LUCIANI, A., LUO, J., LUSSI, Y., MACDOUGALL, A., MADEIRA, F., MAHMOUDY, M., MENCHI, M., MISHRA, A., MOULANG, K., NIGHTINGALE, A., OLIVEIRA, C. S., PUNDIR, S., QI, G., RAJ, S., RICE, D., LOPEZ, M. R., SAIDI, R., SAMPSON, J., SAWFORD, T., SPERETTA, E., TURNER, E., TYAGI, N., VASUDEV, P., VOLYNKIN, V., WARNER, K., WATKINS, X., ZARU, R., ZELLNER, H., BRIDGE, A., POUX, S., REDASCHI, N., AIMO, L., ARGOUD-PUY, G., AUCHINCLOSS, A., AXELSEN, K., BANSAL, P., BARATIN, D., BLATTER, M. C., BOLLEMAN, J., BOUTET, E., BREUZA, L., CASALS-CASAS, C., DE CASTRO, E., ECHIOUKH, K. C., COUDERT, E., CUCHE,

B., DOCHE, M., DORNEVIL, D., ESTREICHER, A., FAMIGLIETTI, M. L., FEUERMANN, M., GASTEIGER, E., GEHANT, S., GERRITSEN, V., GOS, A., GRUAZ-GUMOWSKI, N., HINZ, U., HULO, C., HYKA-NOUSPIKEL, N., JUNGO, F., KELLER, G., KERHORNOU, A., LARA, V., LE MERCIER, P., LIEBERHERR, D., LOMBARDOT, T., MARTIN, X., MASSON, P., MORGAT, A., NETO, T. B., PAESANO, S., PEDRUZZI, I., PILBOUT, S., POURCEL, L., POZZATO, M., PRUESS, M., RIVOIRE, C., SIGRIST, C., SONESSON, K., STUTZ, A., SUNDARAM, S., TOGNOLLI, M., VERBREGUE, L., WU, C. H., ARIGHI, C. N., ARMINSKI, L., CHEN, C., CHEN, Y., GARAVELLI, J. S., HUANG, H., LAIHO, K., MCGARVEY, P., NATALE, D. A., ROSS, K., VINAYAKA, C. R., WANG, Q., WANG, Y., YEH, L. S., ZHANG, J. 2021. UniProt: The universal protein knowledgebase in 2021. In: *Nucleic Acids Research*. 2021, vol. 49, no. D1, pp. D480–D489. ISSN 13624962. Available at: DOI: 10.1093/nar/gkaa1100

BEINART, R. A., ROTTEROVÁ, J., ČEPIČKA, I., GAST, R. J., EDGCOMB, V. P. 2018. The genome of an endosymbiotic methanogen is very similar to those of its free-living relatives. In: *Environmental Microbiology*. 2018, vol. 20, no. 7, pp. 2538–2551. ISSN 14622920. Available at: DOI: 10.1111/1462-2920.14279

BETTS, H. C., PUTTICK, M. N., CLARK, J. W., WILLIAMS, T. A., DONOGHUE, P. C. J., PISANI, D. 2018. Integrated genomic and fossil evidence illuminates life's early evolution and eukaryote origin. In: *Nature Ecology and Evolution*. 2018, vol. 2, no. 10, pp. 1556–1562. ISSN 2397334X. Available at: DOI: 10.1038/s41559-018-0644-x

BEZAWORK-GELETA, A., ROHLENA, J., DONG, L., PACAK, K., NEUZIL, J. 2017. Mitochondrial Complex II: At the Crossroads. In: *Trends in Biochemical Sciences*. 2017, vol. 42, no. 4, pp. 312–325. ISSN 13624326. Available at: DOI: 10.1016/j.tibs.2017.01.003

BODIAN, D. 1936. A new method for staining nerve fibers and nerve endings in mounted paraffin sections. In: *The Anatomical Record*. 1936, vol. 65, no. 1, pp. 89–97. ISSN 0003-276X. Available at: DOI: 10.1002/ar.1090650110

BOLGER, A. M., LOHSE, M., USADEL, B. 2014. Trimmomatic: A flexible trimmer for Illumina sequence data. In: *Bioinformatics*. 2014, vol. 30, no. 15, pp. 2114–2120. ISSN 14602059. Available at: DOI: 10.1093/bioinformatics/btu170

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., ROGER, A. J. 2018. Phylogenomics Places Orphan Protistan Lineages in a Novel Eukaryotic Super-Group. In: *Genome Biology and Evolution*. 2018, vol. 10, no. 2, pp. 427–433. ISSN

17596653. Available at: DOI: 10.1093/gbe/evy014

BROWN, M. W., SHARPE, S. C., SILBERMAN, J. D., HEISS, A. A., LANG, B. F., SIMPSON, A. G. B., ROGER, A. J. 2013. Phylogenomics demonstrates that breviate flagellates are related to opisthokonts and apusomonads. In: *Proceedings of the Royal Society B: Biological Sciences*. 2013, vol. 280, no. 1769. ISSN 14712954. Available at: DOI: 10.1098/rspb.2013.1755

BUCHFINK, B., REUTER, K., DROST, H. G. 2021. Sensitive protein alignments at tree-of-life scale using DIAMOND. In: *Nature Methods*. 2021, vol. 18, no. 4, pp. 366–368. ISSN 15487105. Available at: DOI: 10.1038/s41592-021-01101-x

BUONANNO, F., ORTENZI, C. 2021. Predator-Prey Interactions in Ciliated Protists. In: *Extremophilic Microbes and Metabolites - Diversity, Bioprospecting and Biotechnological Applications*. B.m.: IntechOpen, p. 13. Available at: DOI: 10.5772/intechopen.78622

BURGER, G., GRAY, M. W., FORGET, L., LANG, B. F. 2013. Strikingly bacteria-like and gene-rich mitochondrial genomes throughout jakobid protists. In: *Genome Biology and Evolution*. 2013, vol. 5, no. 2, pp. 418–438. ISSN 17596653. Available at: DOI: 10.1093/gbe/evt008

BURKI, F., ROGER, A. J., BROWN, M. W., SIMPSON, A. G. B. 2020. The New Tree of Eukaryotes. In: *Trends in Ecology and Evolution*. 2020, vol. 35, no. 1, pp. 43–55. ISSN 01695347. Available at: DOI: 10.1016/j.tree.2019.08.008

BURKI, F., SHALCHIAN-TABRIZI, K., MINGE, M., SKJÆVELAND, Å., NIKOLAEV, S. I., JAKOBSEN, K. S., PAWLOWSKI, J. 2007. Phylogenomics reshuffles the eukaryotic supergroups. In: *PLoS ONE*. 2007, vol. 2, no. 8, pp. 1–6. ISSN 19326203. Available at: DOI: 10.1371/journal.pone.0000790

BUSHMANOVA, E., ANTIPOV, D., LAPIDUS, A., PRJIBELSKI, A. D. 2019. RnaSPAdes: A de novo transcriptome assembler and its application to RNA-Seq data. In: *GigaScience*. 2019, vol. 8, no. 9, pp. 1–13. ISSN 2047217X. Available at: DOI: 10.1093/gigascience/giz100

CAMACHO, C., COULOURIS, G., AVAGYAN, V., MA, N., PAPADOPOULOS, J., BEALER, K., MADDEN, T. L. 2009. BLAST+: Architecture and applications. In: *BMC Bioinformatics*. 2009, vol. 10, pp. 1–9. ISSN 14712105. Available at: DOI: 10.1186/1471-2105-10-421

CANTALAPIEDRA, C. P., HERNÁNDEZ-PLAZA, A., LETUNIC, I., HUERTA-CEPAS,

J. 2021. eggNOG-mapper v2: Functional Annotation, Orthology Assignments, and Domain Prediction at the Metagenomic Scale. In: *bioRxiv*. 2021, vol. 5, p. 2021.06.03.446934. DOI: https://doi.org/10.1101/2021.06.03.446934

CAPELLA-GUTIÉRREZ, S., SILLA-MARTÍNEZ, J. M., GABALDÓN, T. 2009. trimAl: A tool for automated alignment trimming in large-scale phylogenetic analyses. In: *Bioinformatics*. 2009, vol. 25, no. 15, pp. 1972–1973. ISSN 13674803. Available at: DOI: 10.1093/bioinformatics/btp348

CAVALIER-SMITH, T. 2002. The phagotrophic origin of eukaryotes and phylogenetic classification on protozoa. In: *International Journal of Systematic and Evolutionary Microbiology*. 2002, vol. 52, no. 2, pp. 297–354. ISSN 14665026. Available at: DOI: 10.1099/00207713-52-2-297

CAVALIER-SMITH, T., CHAO, E. E. Y. 2020. Multidomain ribosomal protein trees and the planctobacterial origin of neomura (eukaryotes, archaebacteria). In: *Protoplasma*. 2020, vol. 257, no. 3, pp. 621–753. ISSN 16156102. Available at: DOI: 10.1007/s00709-019-01442-7

CLAROS, M. G., VINCENS, P. 1996. Computational method to predict mitochondrially imported proteins and their targeting sequences. In: *European Journal of Biochemistry*. 1996, vol. 241, no. 3, pp. 779–786. ISSN 00142956. Available at: DOI: 10.1111/j.1432-1033.1996.00779.x

COCK, P. J. A., ANTAO, T., CHANG, J. T., CHAPMAN, B. A., COX, C. J., DALKE, A., FRIEDBERG, I., HAMELRYCK, T., KAUFF, F., WILCZYNSKI, B., DE HOON, M. J. L. 2009. Biopython: Freely available Python tools for computational molecular biology and bioinformatics. In: *Bioinformatics*. 2009, vol. 25, no. 11, pp. 1422–1423. ISSN 13674803. Available at: DOI: 10.1093/bioinformatics/btp163

COGLIATI, S., LORENZI, I., RIGONI, G., CAICCI, F., SORIANO, M. E. 2018. Regulation of Mitochondrial Electron Transport Chain Assembly. In: *Journal of Molecular Biology*. 2018, vol. 430, no. 24, pp. 4849–4873. ISSN 10898638. Available at: DOI: 10.1016/j.jmb.2018.09.016

CRANFORD-SMITH, T., HUBER, D. 2018. The way is the goal: How SecA transports proteins across the cytoplasmic membrane in bacteria. In: *FEMS Microbiology Letters*. 2018, vol. 365, no. 11, pp. 1–16. ISSN 15746968. Available at: DOI: 10.1093/femsle/fny093

CRISCUOLO, A., GRIBALDO, S. 2010. BMGE (Block Mapping and Gathering with

Entropy): A new software for selection of phylogenetic informative regions from multiple sequence alignments. In: *BMC Evolutionary Biology*. 2010, vol. 10, no. 1. ISSN 14712148. Available at: DOI: 10.1186/1471-2148-10-210

Čepička I (2019) Diverzita protist. Živa 5/2019: 220-223

DACKS, J. B., FIELD, M. C., BUICK, R., EME, L., GRIBALDO, S., ROGER, A. J., BROCHIER-ARMANET, C., DEVOS, D. P. 2016. The changing view of eukaryogenesis - fossils, cells, lineages and how they all come together. In: *Journal of Cell Science*. 2016, vol. 129, no. 20, pp. 3695–3703. ISSN 14779137. Available at: DOI: 10.1242/jcs.178566

DERELLE, R., LANG, B. F. 2012. Rooting the eukaryotic tree with mitochondrial and bacterial proteins. In: *Molecular Biology and Evolution*. 2012, vol. 29, no. 4, pp. 1277–1289. ISSN 07374038. Available at: DOI: 10.1093/molbev/msr295

DERELLE, R., TORRUELLA, G., KLIMEŠ, V., BRINKMANN, H., KIM, E., VLČEK, Č., LANG, B. F., ELIÁŠ, M. 2015. Bacterial proteins pinpoint a single eukaryotic root. In: *Proceedings of the National Academy of Sciences of the United States of America*. 2015, vol. 112, no. 7, pp. E693–E699. ISSN 10916490. Available at: DOI: 10.1073/pnas.1420657112

EKELUND, F. 2002. A study of the soil flagellate Phalansterium solitarium sandon 1924 with preliminary data on its ultrastructure. In: *Protistology*. 2002, vol. 2, no. 3, pp. 152–158. ISSN 1680-0826.

EMANUELSSON, O., NIELSEN, H., BRUNAK, S., VON HEIJNE, G. 2000. Predicting subcellular localization of proteins based on their N-terminal amino acid sequence. In: *Journal of Molecular Biology*. 2000, vol. 300, no. 4, pp. 1005–1016. ISSN 00222836. Available at: DOI: 10.1006/jmbi.2000.3903

FERNANDEZ, D. E. 2018. Two paths diverged in the stroma: targeting to dual SEC translocase systems in chloroplasts. In: *Photosynthesis Research*. 2018, vol. 138, no. 3, pp. 277–287. ISSN 15735079. Available at: DOI: 10.1007/s11120-018-0541-9

FIGUEROA-MARTINEZ, F., JACKSON, C., REYES-PRIETO, A. 2019. Plastid genomes from diverse glaucophyte genera reveal a largely conserved gene content and limited architectural diversity. In: *Genome Biology and Evolution*. 2019, vol. 11, no. 1, pp. 174–188. ISSN 17596653. Available at: DOI: 10.1093/gbe/evy268

FINN, R. D., CLEMENTS, J., EDDY, S. R. 2011. HMMER web server: Interactive sequence similarity searching. In: *Nucleic Acids Research*. 2011, vol. 39, no. SUPPL. 2, pp. 29–37. ISSN 03051048. Available at: DOI: 10.1093/nar/gkr367

FOISSNER, W., BLATTERER, H., FOISSNER, I. 1988. The hemimastigophora (Hemimastix amphikineta nov. gen., nov. spec.), a new protistan phylum from gondwanian soils. In: *European Journal of Protistology*. 1988, vol. 23, no. 4, pp. 361–383. ISSN 09324739. Available at: DOI: 10.1016/S0932-4739(88)80027-0

FOX, G. E., PECHMAN, K. R., WOESE, C. R. 1977. Comparative cataloging of 16S ribosomal ribonucleic acid: molecular approach to procaryotic systematics. In: *International Journal of Systematic Bacteriology*. 1977, vol. 27, no. 1, pp. 44–57. ISSN 00207713. Available at: DOI: 10.1099/00207713-27-1-44

FU, L., NIU, B., ZHU, Z., WU, S., LI, W. 2012. CD-HIT: Accelerated for clustering the next-generation sequencing data. In: *Bioinformatics*. 2012, vol. 28, no. 23, pp. 3150–3152. ISSN 13674803. Available at: DOI: 10.1093/bioinformatics/bts565

FUKASAWA, Y., TSUJI, J., FU, S.-C., TOMII, K., HORTON, P., IMAI, K. 2015. MitoFates: Improved Prediction of Mitochondrial Targeting Sequences and Their Cleavage Sites*. In: *Molecular & Cellular Proteomics*. 2015, vol. 14, no. 4, pp. 1113–1126. ISSN 15359476. Available at: DOI: 10.1074/mcp.M114.043083

GABALDÓN, T., HUYNEN, M. A. 2003. Reconstruction of the proto-mitochondrial metabolism. In: *Science*. 2003, vol. 301, no. 5633, p. 609. ISSN 00368075. Available at: DOI: 10.1126/science.1085463

GABLER, F., NAM, S. Z., TILL, S., MIRDITA, M., STEINEGGER, M., SÖDING, J., LUPAS, A. N., ALVA, V. 2020. Protein Sequence Analysis Using the MPI Bioinformatics Toolkit. In: *Current Protocols in Bioinformatics*. 2020, vol. 72, no. 1, pp. 1–30. ISSN 1934340X. Available at: DOI: 10.1002/cpbi.108

GAGLIARDI, D., STEPIEN, P. P., TEMPERLEY, R. J., LIGHTOWLERS, R. N., CHRZANOWSKA-LIGHTOWLERS, Z. M. A. 2004. Messenger RNA stability in mitochondria: Different means to an end. In: *Trends in Genetics*. 2004, vol. 20, no. 6, pp. 260–267. ISSN 01689525. Available at: DOI: 10.1016/j.tig.2004.04.006

GAWRYLUK, R. M. R., KAMIKAWA, R., STAIRS, C. W., SILBERMAN, J. D., BROWN, M. W., ROGER, A. J. 2016. The Earliest Stages of Mitochondrial Adaptation to Low Oxygen Revealed in a Novel Rhizarian. In: *Current Biology*. 2016, vol. 26, no. 20, pp. 2729–2738. ISSN 09609822. Available at: DOI: 10.1016/j.cub.2016.08.025

GAWRYLUK, R. M. R., STAIRS, C. W. 2021. Diversity of electron transport chains in anaerobic protists. In: *Biochimica et Biophysica Acta - Bioenergetics*. 2021, vol. 1862, no.

1, p. 148334. ISSN 18792650. Available at: DOI: 10.1016/j.bbabio.2020.148334

GAWRYLUK, R. M. R., TIKHONENKOV, D. V., HEHENBERGER, E., HUSNIK, F., MYLNIKOV, A. P., KEELING, P. J. 2019. Non-photosynthetic predators are sister to red algae. In: *Nature*. 2019, vol. 572, no. 7768, pp. 240–243. ISSN 14764687. Available at: DOI: 10.1038/s41586-019-1398-6

GERTS, E. M., YU, Y. K., AGARWALA, R., SCHÄFFER, A. A., ALTSCHUL, S. F. 2006. Composition-based statistics and translated nucleotide searches: Improving the TBLASTN module of BLAST. In: *BMC Biology*. 2006, vol. 4, pp. 1–14. ISSN 17417007. Available at: DOI: 10.1186/1741-7007-4-41

GRABHERR, M. G., HAAS, B. J., YASSOUR, M., LEVIN, J. Z., THOMPSON, D. A., AMIT, I., ADICONIS, X., FAN, L., RAYCHOWDHURY, R., ZENG, Q., CHEN, Z., MAUCELI, E., HACOHEN, N., GNIRKE, A., RHIND, N., DI PALMA, F., BIRREN, B. W., NUSBAUM, C., LINDBLAD-TOH, K., FRIEDMAN, N., REGEV, A. 2011. Full-length transcriptome assembly from RNA-Seq data without a reference genome. In: *Nature Biotechnology*. 2011, vol. 29, no. 7, pp. 644–652. ISSN 10870156. Available at: DOI: 10.1038/nbt.1883

GRAY, M. W. 2015. Mosaic nature of the mitochondrial proteome: Implications for the origin and evolution of mitochondria. In: *Proceedings of the National Academy of Sciences of the United States of America*. 2015, vol. 112, no. 33, pp. 10133–10138. ISSN 10916490. Available at: DOI: 10.1073/pnas.1421379112

GRAY, M. W., BURGER, G., DERELLE, R., KLIMEŠ, V., LEGER, M. M., SARRASIN, M., VLČEK, Č., ROGER, A. J., ELIÁŠ, M., 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*. ISBN 1291502007416. Available at: DOI: 10.1186/s12915-020-0741-6

GRAY, M. W., LANG, B. F., BURGER, G. 2004. Mitochondria of protists. In: *Annual Review of Genetics*. 2004, vol. 38, pp. 477–524. ISSN 00664197. Available at: DOI: 10.1146/annurev.genet.37.110801.142526

GREINER, S., LEHWARK, P., BOCK, R. 2019. OrganellarGenomeDRAW (OGDRAW) version 1.3.1: Expanded toolkit for the graphical visualization of organellar genomes. In: *Nucleic Acids Research*. 2019, vol. 47, no. W1, pp. W59–W64. ISSN 13624962. Available at: DOI: 10.1093/nar/gkz238

HADARIOVÁ, L., VESTEG, M., HAMPL, V., KRAJČOVIČ, J. 2018. Reductive evolution of chloroplasts in non-photosynthetic plants, algae and protists. In: *Current Genetics*. 2018, vol. 64, no. 2, pp. 365–387. ISSN 14320983. Available at: DOI: 10.1007/s00294-017-0761-0

HAFT, D. H., SELENGUT, J. D., RICHTER, R. A., HARKINS, D., BASU, M. K., BECK, E. 2013. TIGRFAMs and genome properties in 2013. In: *Nucleic Acids Research*. 2013, vol. 41, no. D1, pp. 387–395. ISSN 03051048. Available at: DOI: 10.1093/nar/gks1234

HAJNSDORF, E., KABERDIN, V. R. 2018. RNA polyadenylation and its consequences in prokaryotes. In: *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2018, vol. 373, no. 1762, pp. 1–9. ISSN 14712970. Available at: DOI: 10.1098/rstb.2018.0166

HALL, T. 2004. BioEdit version 7.0. 0. In: *Distributed by the author, website: www.mbio.ncsu.edu/BioEdit/bioedit.html*. 2004.

HAMPL, V. 2017. Preaxostyla. In: Slamovits C ARCHIBALD J, SIMPSON A, ed. *Handbook of the Protists*. Cham (Switzerland): Springer International Publishing, p. 1139–1174.

HAMPL, V., ČEPIČKA, I., ELIÁŠ, M. 2019. Was the Mitochondrion Necessary to Start Eukaryogenesis? In: *Trends in Microbiology*. 2019, vol. 27, no. 2, pp. 96–104. ISSN 18784380. Available at: DOI: 10.1016/j.tim.2018.10.005

HAUSMANN, K. 1978. Extrusive Organelles in Protists. In: *International Review of Cytology*. 1978, vol. 52, no. C, pp. 197–276. ISSN 00747696. Available at: DOI: 10.1016/S0074-7696(08)60757-3

HE, D., FIZ-PALACIOS, O., FU, C. J., TSAI, C. C., BALDAUF, S. L. 2014. An alternative root for the eukaryote tree of life. In: *Current Biology*. 2014, vol. 24, no. 4, pp. 465–470. ISSN 09609822. Available at: DOI: 10.1016/j.cub.2014.01.036

HEISS, A. A., KOLISKO, M., EKELUND, F., BROWN, M. W., ROGER, A. J., SIMPSON, A. G. B. 2018. Combined morphological and phylogenomic re-examination of malawimonads, a critical taxon for inferring the evolutionary history of eukaryotes. In: *Royal Society Open Science*. 2018, vol. 5, no. 4. ISSN 20545703. Available at: DOI: 10.1098/rsos.171707

HOANG, D. T., CHERNOMOR, O., VON HAESELER, A., MINH, B. Q., VINH, L. S. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. In: *Molecular Biology and*

Evolution. 2018, vol. 35, no. 2, pp. 518–522. ISSN 15371719. Available at: DOI: 10.1093/molbev/msx281

HOLZMANN, M., PAWLOWSKI, J. 2017. An updated classification of rotaliid foraminifera based on ribosomal DNA phylogeny. In: *Marine Micropaleontology*. 2017, vol. 132, no. April, pp. 18–34. ISSN 03778398. Available at: DOI: 10.1016/j.marmicro.2017.04.002

HORTON, P., PARK, K. J., OBAYASHI, T., FUJITA, N., HARADA, H., ADAMS-COLLIER, C. J., NAKAI, K. 2007. WoLF PSORT: Protein localization predictor. In: *Nucleic Acids Research*. 2007, vol. 35, no. SUPPL.2, pp. 585–587. ISSN 03051048. Available at: DOI: 10.1093/nar/gkm259

HORVÁTHOVÁ, L., ŽÁRSKÝ, V., PÁNEK, T., DERELLE, R., PYRIH, J., MOTYČKOVÁ, A., KLÁPŠŤOVÁ, V., VINOPALOVÁ, M., MARKOVÁ, L., VOLEMAN, L., KLIMEŠ, V., PETRŮ, M., VAITOVÁ, Z., ČEPIČKA, I., HRYZÁKOVÁ, K., HARANT, K., GRAY, M. W., CHAMI, M., GUILVOUT, I., FRANCETIC, O., FRANZ LANG, B., VLČEK, Č., TSAOUSIS, A. D., ELIÁŠ, M., DOLEŽAL, P. 2021. Analysis of diverse eukaryotes suggests the existence of an ancestral mitochondrial apparatus derived from the bacterial type II secretion system. In: *Nature Communications*. 2021, vol. 12, no. 1, pp. 1–18. ISSN 20411723. Available at: DOI: 10.1038/s41467-021-23046-7

HOWE, A. T., BASS, D., CHAO, E. E., CAVALIER-SMITH, T. 2011. New Genera, Species, and Improved Phylogeny of Glissomonadida (Cercozoa). In: *Protist*. 2011, vol. 162, no. 5, pp. 710–722. ISSN 14344610. Available at: DOI: 10.1016/j.protis.2011.06.002

HUERTA-CEPAS, J., SZKLARCZYK, D., HELLER, D., HERNÁNDEZ-PLAZA, A., FORSLUND, S. K., COOK, H., MENDE, D. R., LETUNIC, I., RATTEI, T., JENSEN, L. J., VON MERING, C., BORK, P. 2019. EggNOG 5.0: A hierarchical, functionally and phylogenetically annotated orthology resource based on 5090 organisms and 2502 viruses. In: *Nucleic Acids Research*. 2019, vol. 47, no. D1, pp. D309–D314. ISSN 13624962. Available at: DOI: 10.1093/nar/gky1085

IRISARRI, I., STRASSERT, J. F. H., BURKI, F. 2021. Phylogenomic Insights into the Origin of Primary Plastids. In: *Systematic Biology*. 2021, vol. 0, no. 0, pp. 1–16. ISSN 1063-5157. Available at: DOI: 10.1093/sysbio/syab036

JANOUŠKOVEC, J., TIKHONENKOV, D. V., BURKI, F., HOWE, A. T., ROHWER, F. L., MYLNIKOV, A. P., KEELING, P. J. 2017. A New Lineage of Eukaryotes Illuminates

Early Mitochondrial Genome Reduction. In: *Current Biology*. 2017, vol. 27, no. 23, pp. 3717-3724.e5. ISSN 09609822. Available at: DOI: 10.1016/j.cub.2017.10.051

JARDIM-MESSEDER, D., CABREIRA-CAGLIARI, C., RAUBER, R., TURCHETTO-ZOLET, A. C., MARGIS, R., MARGIS-PINHEIRO, M. 2017. Fumarate reductase superfamily: A diverse group of enzymes whose evolution is correlated to the establishment of different metabolic pathways. In: *Mitochondrion*. 2017, vol. 34, pp. 56–66. ISSN 18728278. Available at: DOI: 10.1016/j.mito.2017.01.002

JONES, P., BINNS, D., CHANG, H. Y., FRASER, M., LI, W., MCANULLA, C., MCWILLIAM, H., MASLEN, J., MITCHELL, A., NUKA, G., PESSEAT, S., QUINN, A. F., SANGRADOR-VEGAS, A., SCHEREMETJEW, M., YONG, S. Y., LOPEZ, R., HUNTER, S. 2014. InterProScan 5: Genome-scale protein function classification. In: *Bioinformatics*. 2014, vol. 30, no. 9, pp. 1236–1240. ISSN 14602059. Available at: DOI: 10.1093/bioinformatics/btu031

JUERGENS, H., MIELGO-GÓMEZ, Á., GODOY-HERNÁNDEZ, A., TER HORST, J., NIJENHUIS, J. M., MCMILLAN, D. G. G., MANS, R. 2021. Physiological relevance, localization and substrate specificity of the alternative (type II) mitochondrial NADH dehydrogenases of Ogataea parapolymorpha. In: *bioRxiv*. 2021, p. 2021.04.28.441406. Available at: DOI: https://doi.org/10.1101/2021.04.28.441406

KALYAANAMOORTHY, S., MINH, B. Q., WONG, T. K. F., VON HAESELER, A., JERMIIN, L. S. 2017. ModelFinder: Fast model selection for accurate phylogenetic estimates. In: *Nature Methods*. 2017, vol. 14, no. 6, pp. 587–589. ISSN 15487105. Available at: DOI: 10.1038/nmeth.4285

KANEHISA, M., SATO, Y. 2020. KEGG Mapper for inferring cellular functions from protein sequences. In: *Protein Science*. 2020, vol. 29, no. 1, pp. 28–35. ISSN 1469896X. Available at: DOI: 10.1002/pro.3711

KANEHISA, M., SATO, Y., MORISHIMA, K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for Functional Characterization of Genome and Metagenome Sequences. In: *Journal of Molecular Biology*. 2016, vol. 428, no. 4, pp. 726–731. ISSN 10898638. Available at: DOI: 10.1016/j.jmb.2015.11.006

KANG, S., TICE, A. K., SPIEGEL, F. W., SILBERMAN, J. D., PÁNEK, T., ČEPIČKA, I., KOSTKA, M., KOSAKYAN, A., ALCÂNTARA, D. M. C. C., ROGER, A. J., SHADWICK, L. L., SMIRNOV, A., KUDRYAVTSEV, A., LAHR, D. J. G. G., BROWN,

M. W. 2017. Between a Pod and a Hard Test: The Deep Evolution of Amoebae. In: *Molecular Biology and Evolution*. 2017, vol. 34, no. 9, pp. 2258–2270. ISSN 0737-4038. Available at: DOI: 10.1093/molbev/msx162

KARNKOWSKA, A., TREITLI, S. C., BRZOŇ, O., NOVÁK, L., VACEK, V., SOUKAL, P., BARLOW, L. D., HERMAN, E. K., PIPALIYA, S. V., PÁNEK, T., ŽIHALA, D., PETRŽELKOVÁ, R., BUTENKO, A., EME, L., STAIRS, C. W., ROGER, A. J., ELIÁŠ, M., DACKS, J. B., HAMPL, V., BATTISTUZZI, F. U. 2019. The Oxymonad Genome Displays Canonical Eukaryotic Complexity in the Absence of a Mitochondrion. In: *Molecular Biology and Evolution*. 2019, vol. 36, no. 10, pp. 2292–2312. ISSN 15371719. Available at: DOI: 10.1093/molbev/msz147

KARNKOWSKA, A., VACEK, V., ZUBÁČOVÁ, Z., TREITLI, S. C., PETRŽELKOVÁ, R., EME, L., NOVÁK, L., ŽÁRSKÝ, V., BARLOW, L. D., HERMAN, E. K., SOUKAL, P., HROUDOVÁ, M., DOLEŽAL, P., STAIRS, C. W., ROGER, A. J., ELIÁŠ, M., DACKS, J. B., VLČEK, Č., HAMPL, V. 2016. A eukaryote without a mitochondrial organelle. In: *Current Biology*. 2016, vol. 26, no. 10, pp. 1274–1284. ISSN 09609822. Available at: DOI: 10.1016/j.cub.2016.03.053

KATOH, K., ROZEWICKI, J., YAMADA, K. D. 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. In: *Briefings in Bioinformatics*. 2019, vol. 20, no. 4, pp. 1160–1166. ISSN 1467-5463. Available at: DOI: 10.1093/bib/bbx108

KATOH, K., STANDLEY, D. M. 2013. MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. In: *Molecular Biology and Evolution*. 2013, vol. 30, no. 4, pp. 772–780. ISSN 07374038. Available at: DOI: 10.1093/molbev/mst010

KAUR, B., ZÁHONOVÁ, K., VALACH, M., FAKTOROVÁ, D., PROKOPCHUK, G., BURGER, G., LUKEŠ, J. 2020. Gene fragmentation and RNA editing without borders: Eccentric mitochondrial genomes of diplonemids. In: *Nucleic Acids Research*. 2020, vol. 48, no. 5, pp. 2694–2708. ISSN 13624962. Available at: DOI: 10.1093/nar/gkz1215

KEELING, P. J., BURKI, F. 2019. Progress towards the Tree of Eukaryotes. In: *Current Biology*. 2019, vol. 29, no. 16, pp. R808–R817. ISSN 09609822. Available at: DOI: 10.1016/j.cub.2019.07.031

KIM, D., PAGGI, J. M., PARK, C., BENNETT, C., SALZBERG, S. L. 2019. Graph-based

genome alignment and genotyping with HISAT2 and HISAT-genotype. In: *Nature Biotechnology*. 2019, vol. 37, no. 8, pp. 907–915. ISSN 15461696. Available at: DOI: 10.1038/s41587-019-0201-4

KIM, J. I., JEONG, M., ARCHIBALD, J. M., SHIN, W. 2020. Comparative Plastid Genomics of Non-Photosynthetic Chrysophytes: Genome Reduction and Compaction. In: *Frontiers in Plant Science*. 2020, vol. 11, no. September. ISSN 1664462X. Available at: DOI: 10.3389/fpls.2020.572703

KIM, J. I., MOORE, C. E., ARCHIBALD, J. M., BHATTACHARYA, D., YI, G., YOON, H. S., SHIN, W. 2017. Evolutionary dynamics of cryptophyte plastid genomes. In: *Genome Biology and Evolution*. 2017, vol. 9, no. 7, pp. 1859–1872. ISSN 17596653. Available at: DOI: 10.1093/gbe/evx123

KLEBS, G. A. 1892. Flagellatenstudien. Theil II. In: *Zeitschrift für wissenschaftliche Zoologie*. 1892, vol. 55, pp. 352–445.

KOUMANDOU, V. L., WICKSTEAD, B., GINGER, M. L., VAN DER GIEZEN, M., DACKS, J. B., FIELD, M. C. 2013. Molecular paleontology and complexity in the last eukaryotic common ancestor. In: *Critical Reviews in Biochemistry and Molecular Biology*. 2013, vol. 48, no. 4, pp. 373–396. ISSN 10409238. Available at: DOI: 10.3109/10409238.2013.821444

KOYAMA, Y., TAKIMOTO, K., KOJIMA, A., ASAI, K., MATSUOKA, S., MITSUI, T., MATSUMOTO, K., HARA, H., OHTA, N. 2011. Characterization of the nuclear- and plastid-encoded seca-homologous genes in the unicellular red alga cyanidioschyzon merolae. In: *Bioscience, Biotechnology and Biochemistry*. 2011, vol. 75, no. 10, pp. 2073–2078. ISSN 09168451. Available at: DOI: 10.1271/bbb.110338

KUGRENS, P., LEE, R. E., CORLISS, J. O. 1994. Ultrastructure, biogenesis, and functions of extrusive organelles in selected non-ciliate protists. In: *Protoplasma*. 1994, vol. 181, nos. 1–4, pp. 164–190. ISSN 0033183X. Available at: DOI: 10.1007/BF01666394

LAHR, D. J. G., KOSAKYAN, A., LARA, E., MITCHELL, E. A. D., MORAIS, L., PORFIRIO-SOUSA, A. L., RIBEIRO, G. M., TICE, A. K., PÁNEK, T., KANG, S., BROWN, M. W. 2019. Phylogenomics and Morphological Reconstruction of Arcellinida Testate Amoebae Highlight Diversity of Microbial Eukaryotes in the Neoproterozoic. In: *Current Biology*. 2019, vol. 29, no. 6, pp. 991-1001.e3. ISSN 09609822. Available at: DOI: 10.1016/j.cub.2019.01.078

LAX, G., EGLIT, Y., EME, L., BERTRAND, E. M., ROGER, A. J., SIMPSON, A. G. B. 2018. Hemimastigophora is a novel supra-kingdom-level lineage of eukaryotes. In: *Nature*. 2018, vol. 564, no. 7736, pp. 410–414. ISSN 14764687. Available at: DOI: 10.1038/s41586-018-0708-8

LE, S. Q., DANG, C. C., GASCUEL, O. 2012. Modeling protein evolution with several amino acid replacement matrices depending on site rates. In: *Molecular Biology and Evolution*. 2012, vol. 29, no. 10, pp. 2921–2936. ISSN 07374038. Available at: DOI: 10.1093/molbev/mss112

LEGER, M. M., KOLISKO, M., KAMIKAWA, R., STAIRS, C. W., KUME, K., ČEPIČKA, I., SILBERMAN, J. D., ANDERSSON, J. O., XU, F., YABUKI, A., EME, L., ZHANG, Q., TAKISHITA, K., INAGAKI, Y., SIMPSON, A. G. B., HASHIMOTO, T., ROGER, A. J. 2017. Organelles that illuminate the origins of Trichomonas hydrogenosomes and Giardia mitosomes. In: *Nature Ecology and Evolution*. 2017, vol. 1, no. 4, pp. 1–7. ISSN 2397334X. Available at: DOI: 10.1038/s41559-017-0092

LEWIS, W. H., LIND, A. E., SENDRA, K. M., ONSBRING, H., WILLIAMS, T. A., ESTEBAN, G. F., HIRT, R. P., ETTEMA, T. J. G., EMBLEY, T. M. 2020. Convergent Evolution of Hydrogenosomes from Mitochondria by Gene Transfer and Loss. In: *Molecular Biology and Evolution*. 2020, vol. 37, no. 2, pp. 524–539. ISSN 15371719. Available at: DOI: 10.1093/molbev/msz239

LUKEŠ, J., WHEELER, R., JIRSOVÁ, D., DAVID, V., ARCHIBALD, J. M. 2018. Massive mitochondrial DNA content in diplonemid and kinetoplastid protists. In: *IUBMB Life*. 2018, vol. 70, no. 12, pp. 1267–1274. ISSN 15216551. Available at: DOI: 10.1002/iub.1894

MANNI, M., BERKELEY, M. R., SEPPEY, M., SIMAO, F. A., 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. 2021. ArXiv ID 2106.11799

MARANDE, W., LÓPEZ-GARCÍA, P., MOREIRA, D. 2009. Eukaryotic diversity and phylogeny using small- and large-subunit ribosomal RNA genes from environmental samples. In: *Environmental Microbiology*. 2009, vol. 11, no. 12, pp. 3179–3188. ISSN 14622912. Available at: DOI: 10.1111/j.1462-2920.2009.02023.x

MARTIN, W. F. 2018. Eukaryote lateral gene transfer is Lamarckian. In: *Nature Ecology and Evolution*. 2018, vol. 2, no. 5, p. 754. ISSN 2397334X. Available at: DOI:

10.1038/s41559-018-0521-7

MARTIN, W. F., TIELENS, A. G. M., MENTEL, M., GARG, S. G., GOULD, S. B. 2017. The Physiology of Phagocytosis in the Context of Mitochondrial Origin. In: *Microbiology and Molecular Biology Reviews*. 2017, vol. 81, no. 3. ISSN 1092-2172. Available at: DOI: 10.1128/mmbr.00008-17

MEDLIN, L., ELWOOD, H. J., STICKEL, S., SOGIN, M. L. 1988. The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. In: *Gene*. 1988, vol. 71, no. 2, pp. 491–499. ISSN 03781119. Available at: DOI: 10.1016/0378-1119(88)90066-2

MINH, B. Q., SCHMIDT, H. A., CHERNOMOR, O., SCHREMPF, D., WOODHAMS, M. D., VON HAESELER, A., LANFEAR, R., TEELING, E. 2020. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. In: *Molecular Biology and Evolution*. 2020, vol. 37, no. 5, pp. 1530–1534. ISSN 15371719. Available at: DOI: 10.1093/molbev/msaa015

MISTRY, J., CHUGURANSKY, S., WILLIAMS, L., QURESHI, M., SALAZAR, G. A., SONNHAMMER, E. L. L., TOSATTO, S. C. E., PALADIN, L., RAJ, S., RICHARDSON, L. J., FINN, R. D., BATEMAN, A. 2021. Pfam: The protein families database in 2021. In: *Nucleic Acids Research*. 2021, vol. 49, no. D1, pp. D412–D419. ISSN 13624962. Available at: DOI: 10.1093/nar/gkaa913

MOOSAVI, B., BERRY, E. A., ZHU, X. L., YANG, W. C., YANG, G. F. 2019. The assembly of succinate dehydrogenase: a key enzyme in bioenergetics. In: *Cellular and Molecular Life Sciences*. 2019, vol. 76, no. 20, pp. 4023–4042. ISSN 14209071. Available at: DOI: 10.1007/s00018-019-03200-7

MORALES, J., MOGI, T., MINEKI, S., TAKASHIMA, E., MINEKI, R., HIRAWAKE, H., SAKAMOTO, K., OMURA, S., KITA, K. 2009. Novel mitochondrial complex II isolated from trypanosoma cruzi is composed of 12 peptides including a heterodimeric ip subunit. In: *Journal of Biological Chemistry*. 2009, vol. 284, no. 11, pp. 7255–7263. ISSN 1083351X. Available at: DOI: 10.1074/jbc.M806623200

MORLAN, J. D., QU, K., SINICROPI, D. V. 2012. Selective depletion of rRNA enables whole transcriptome profiling of archival fixed tissue. In: *PLoS ONE*. 2012, vol. 7, no. 8, pp. 1–8. ISSN 19326203. Available at: DOI: 10.1371/journal.pone.0042882

MÜLLER, M., MENTEL, M., VAN HELLEMOND, J. J., HENZE, K., WOEHLE, C., GOULD, S. B., YU, R.-Y., VAN DER GIEZEN, M., TIELENS, A. G. M., MARTIN, W. F.

2012. Biochemistry and Evolution of Anaerobic Energy Metabolism in Eukaryotes. In: *Microbiology and Molecular Biology Reviews*. 2012, vol. 76, no. 2, pp. 444–495. ISSN 1092-2172. Available at: DOI: 10.1128/mmbr.05024-11

MYL'NIKOVA, A. A., MYL'NIKOV, A. P. 2011. Ultrastructure of the marine predatory flagellate Metromonas simplex Larsen et Patterson, 1990 (Cercozoa). In: *Inland Water Biology*. 2011, vol. 4, no. 2, pp. 105–110. ISSN 19950829. Available at: DOI: 10.1134/S1995082911020155

MYLNIKOV, A. P., MYLNIKOV, A. A. 2014. Structure of the flagellar apparatus of the bacterivorous flagellate Histiona aroides Pascher, 1943 (Jakobida, Excavata). In: *Inland Water Biology*. 2014, vol. 7, no. 4, pp. 331–337. ISSN 19950837. Available at: DOI: 10.1134/S1995082914040130

NGUYEN, D. H., NGUYEN, C. H., MAMITSUKA, H. 2019. Recent advances and prospects of computational methods for metabolite identification: A review with emphasis on machine learning approaches. In: *Briefings in Bioinformatics*. 2019, vol. 20, no. 6, pp. 2028–2043. ISSN 14774054. Available at: DOI: 10.1093/bib/bby066

NGUYEN, L. T., SCHMIDT, H. A., VON HAESELER, A., MINH, B. Q. 2015. IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. In: *Molecular Biology and Evolution*. 2015, vol. 32, no. 1, pp. 268–274. ISSN 15371719. Available at: DOI: 10.1093/molbev/msu300

NIE, D. 1950. Morphology and taxonomy of the intestinal protozoa of the guinea-pig, cavia porcella. In: *Journal of Morphology*. 1950, vol. 86, no. 3, pp. 381–493. ISSN 0362-2525. Available at: DOI: 10.1002/jmor.1050860302

NÝVLTOVÁ, E., STAIRS, C. W., HRDÝ, I., RÍDL, J., MACH, J., PAЧES, J., ROGER, A. J., TACHEZY, J. 2015. Lateral gene transfer and gene duplication played a key role in the evolution of mastigamoeba balamuthi hydrogenosomes. In: *Molecular Biology and Evolution*. 2015, vol. 32, no. 4, pp. 1039–1055. ISSN 15371719. Available at: DOI: 10.1093/molbev/msu408

O'LEARY, N. A., WRIGHT, M. W., BRISTER, J. R., CIUFO, S., HADDAD, D., MCVEIGH, R., RAJPUT, B., ROBBERTSE, B., SMITH-WHITE, B., AKO-ADJEI, D., ASTASHYN, A., BADRETDIN, A., BAO, Y., BLINKOVA, O., BROVER, V., CHETVERNIN, V., CHOI, J., COX, E., ERMOLAEVA, O., FARRELL, C. M., GOLDFARB, T., GUPTA, T., HAFT, D., HATCHER, E., HLAVINA, W., JOARDAR, V.

S., KODALI, V. K., LI, W., MAGLOTT, D., MASTERSON, P., MCGARVEY, K. M., MURPHY, M. R., O'NEILL, K., PUJAR, S., RANGWALA, S. H., RAUSCH, D., RIDDICK, L. D., SCHOCH, C., SHKEDA, A., STORZ, S. S., SUN, H., THIBAUD-NISSEN, F., TOLSTOY, I., TULLY, R. E., VATSAN, A. R., WALLIN, C., WEBB, D., WU, W., LANDRUM, M. J., KIMCHI, A., TATUSOVA, T., DICUCCIO, M., KITTS, P., MURPHY, T. D., PRUITT, K. D. 2016. Reference sequence (RefSeq) database at NCBI: Current status, taxonomic expansion, and functional annotation. In: *Nucleic Acids Research*. 2016, vol. 44, no. D1, pp. D733–D745. ISSN 13624962. Available at: DOI: 10.1093/nar/gkv1189

OBORNÍK, M., VANCOVÁ, M., LAI, D. H., JANOUŠKOVEC, J., KEELING, P. J., LUKEŠ, J. 2011. Morphology and ultrastructure of multiple life cycle stages of the photosynthetic relative of apicomplexa, Chromera velia. In: *Protist*. 2011, vol. 162, no. 1, pp. 115–130. ISSN 14344610. Available at: DOI: 10.1016/j.protis.2010.02.004

PÁNEK, T., SIMPSON, A. G. B., HAMPL, V., ČEPIČKA, I. 2014. Creneis carolina gen. et sp. nov. (Heterolobosea), a Novel Marine Anaerobic Protist with Strikingly Derived Morphology and Life Cycle. In: *Protist*. 2014, vol. 165, no. 4, pp. 542–567. ISSN 16180941. Available at: DOI: 10.1016/j.protis.2014.05.005

PATTERSON, D. J. 1990. Jakoba libera (Ruinen, 1938), a heterotrophic flagellate from deep oceanic sediments. In: *Journal of the Marine Biological Association of the United Kingdom*. 1990, vol. 70, no. 2, pp. 381–393. ISSN 0025-3154. Available at: DOI: 10.1017/S0025315400035487

PAWLOWSKI, J., AUDIC, S., ADL, S., BASS, D., BELBAHRI, L., BERNEY, C., BOWSER, S. S., CEPICKA, I., DECELLE, J., DUNTHORN, M., FIORE-DONNO, A. M., GILE, G. H., HOLZMANN, M., JAHN, R., JIRKŮ, M., KEELING, P. J., KOSTKA, M., KUDRYAVTSEV, A., LARA, E., LUKEŠ, J., MANN, D. G., MITCHELL, E. A. D., NITSCHE, F., ROMERALO, M., SAUNDERS, G. W., SIMPSON, A. G. B., SMIRNOV, A. V., SPOUGE, J. L., STERN, R. F., STOECK, T., ZIMMERMANN, J., SCHINDEL, D., DE VARGAS, C. 2012. CBOL Protist Working Group: Barcoding Eukaryotic Richness beyond the Animal, Plant, and Fungal Kingdoms. In: *PLoS Biology*. 2012, vol. 10, no. 11. ISSN 15449173. Available at: DOI: 10.1371/journal.pbio.1001419

PETRU, M., WIDEMAN, J., MOORE, K., ALCOCK, F., PALMER, T., DOLEŽAL, P. 2018. Evolution of mitochondrial TAT translocases illustrates the loss of bacterial protein transport machines in mitochondria. In: *BMC Biology*. 2018, vol. 16, no. 1, pp. 1–14. ISSN

17417007. Available at: DOI: 10.1186/s12915-018-0607-3

PITTIS, A. A., GABALDÓN, T. 2016. Late acquisition of mitochondria by a host with chimaeric prokaryotic ancestry. In: *Nature*. 2016, vol. 531, no. 7592, pp. 101–104. ISSN 14764687. Available at: DOI: 10.1038/nature16941

PLATTNER, H. 2017. Trichocysts—Paramecium's Projectile-like Secretory Organelles: Reappraisal of their Biogenesis, Composition, Intracellular Transport, and Possible Functions. In: *Journal of Eukaryotic Microbiology*. 2017, vol. 64, no. 1, pp. 106–133. ISSN 15507408. Available at: DOI: 10.1111/jeu.12332

PRICE, M. N., DEHAL, P. S., ARKIN, A. P. 2010. FastTree 2 - Approximately maximum-likelihood trees for large alignments. In: *PLoS ONE*. 2010, vol. 5, no. 3. ISSN 19326203. Available at: DOI: 10.1371/journal.pone.0009490

PYRIH, J., PÁNEK, T., DURANTE, I. M., RAŠKOVÁ, V., CIMRHANZLOVÁ, K., KRIEGOVÁ, E., TSAOUSIS, A. D., ELIÁŠ, M., LUKEŠ, J. 2021. Vestiges of the Bacterial Signal Recognition Particle-Based Protein Targeting in Mitochondria. In: *Molecular Biology and Evolution*. 2021. ISSN 0737-4038. Available at: DOI: 10.1093/molbev/msab090

RHIEL, E. 2017. On the extrusomes of Oxyrrhis marina (Dinophyceae). In: *Protoplasma*. 2017, vol. 254, no. 2, pp. 901–909. ISSN 0033183X. Available at: DOI: 10.1007/s00709-016-0999-2

RIO, D. C., ARES, M., HANNON, G. J., NILSEN, T. W. 2010. Purification of RNA using TRIzol (TRI Reagent). In: *Cold Spring Harbor Protocols*. 2010, vol. 5, no. 6, pp. 1–4. ISSN 15596095. Available at: DOI: 10.1101/pdb.prot5439

ROGER, A. J., MUÑOZ-GÓMEZ, S. A., KAMIKAWA, R. 2017. The Origin and Diversification of Mitochondria. In: *Current Biology*. 2017, vol. 27, no. 21, pp. R1177–R1192. ISSN 09609822. Available at: DOI: 10.1016/j.cub.2017.09.015

ROGER, A. J., SUSKO, E., LEGER, M. M. 2021. Evolution: Reconstructing the Timeline of Eukaryogenesis. In: *Current Biology*. 2021, vol. 31, no. 4, pp. R193–R196. ISSN 18790445. Available at: DOI: 10.1016/j.cub.2020.12.035

ROSATI, G., MODEO, L. 2003. Extrusomes in Ciliates: Diversification, Distribution, and Phylogenetic Implications. In: *The Journal of Eukaryotic Microbiology*. 2003, vol. 50, no. 6, pp. 383–402. ISSN 1066-5234. Available at: DOI: 10.1111/j.1550-7408.2003.tb00260.x

SAKAGUCHI, M., SUZAKI, T., KAMAL KHAN, S. M. M., HAUSMANN, K. 2002. Food capture by kinetocysts in the heliozoon Raphidiophrys contractilis. In: *European Journal of*

Protistology. 2002, vol. 37, no. 4, pp. 453–458. ISSN 09324739. Available at: DOI: 10.1078/0932-4739-00847

SAVOJARDO, C., BRUCIAFERRI, N., TARTARI, G., MARTELLI, P. L., CASADIO, R. 2020. DeepMito: Accurate prediction of protein sub-mitochondrial localization using convolutional neural networks. In: *Bioinformatics*. 2020, vol. 36, no. 1, pp. 56–64. ISSN 14602059. Available at: DOI: 10.1093/bioinformatics/btz512

SAVOJARDO, C., MARTELLI, P. L., FARISELLI, P., CASADIO, R. 2015. TPpred3 detects and discriminates mitochondrial and chloroplastic targeting peptides in eukaryotic proteins. In: *Bioinformatics*. 2015, vol. 31, no. 20, pp. 3269–3275. ISSN 14602059. Available at: DOI: 10.1093/bioinformatics/btv367

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., BURKI, F. 2021. Picozoa are archaeplastids without plastid. In: *bioRxiv*. 2021, pp. 1–17. DOI: https://doi.org/10.1101/2021.04.14.439778

SCHUT, G. J., ADAMS, M. W. W. 2009. The iron-hydrogenase of Thermotoga maritima utilizes ferredoxin and NADH synergistically: A new perspective on anaerobic hydrogen production. In: *Journal of Bacteriology*. 2009, vol. 191, no. 13, pp. 4451–4457. ISSN 00219193. Available at: DOI: 10.1128/JB.01582-08

SPANG, A., SAW, J. H., JØRGENSEN, S. L., ZAREMBA-NIEDZWIEDZKA, K., MARTIJN, J., LIND, A. E., VAN EIJK, R., SCHLEPER, C., GUY, L., ETTEMA, T. J. G. 2015. Complex archaea that bridge the gap between prokaryotes and eukaryotes. In: *Nature*. 2015, vol. 521, no. 7551, pp. 173–179. ISSN 14764687. Available at: DOI: 10.1038/nature14447

STAIRS, C. W., LEGER, M. M., ROGER, A. J. 2015. Diversity and origins of anaerobic metabolism in mitochondria and related organelles. In: *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2015, vol. 370, no. 1678. ISSN 14712970. Available at: DOI: 10.1098/rstb.2014.0326

STAMATAKIS, A. 2006. RAxML-VI-HPC: Maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. In: *Bioinformatics*. 2006, vol. 22, no. 21, pp. 2688–2690. ISSN 13674803. Available at: DOI: 10.1093/bioinformatics/btl446

STEINER, J. M., BHATTACHARYA, D., LÖFFELHARDT, W. 2012. Conservative sorting in the muroplasts of Cyanophora paradoxa: A reevaluation based on the completed

genome sequence. In: *Symbiosis*. 2012, vol. 58, nos. 1–3, pp. 127–133. ISSN 03345114. Available at: DOI: 10.1007/s13199-012-0203-1

STRASSERT, J. F. H., IRISARRI, I., WILLIAMS, T. A., BURKI, F. 2021. A molecular timescale for eukaryote evolution with implications for the origin of red algal-derived plastids. In: *Nature Communications*. 2021, vol. 12, no. 1. ISSN 20411723. Available at: DOI: 10.1038/s41467-021-22044-z

STRASSERT, J. F. H., JAMY, M., MYLNIKOV, A. P., TIKHONENKOV, D. V., BURKI, F. 2019. New phylogenomic analysis of the enigmatic phylum Telonemia further resolves the eukaryote tree of life. In: *Molecular Biology and Evolution*. 2019, vol. 36, no. 4, pp. 757–765. ISSN 15371719. Available at: DOI: 10.1093/molbev/msz012

SUN, C., RUSCH, S. L., KIM, J., KENDALL, D. A. 2007. Chloroplast SecA and Escherichia coli SecA have distinct lipid and signal peptide preferences. In: *Journal of Bacteriology*. 2007, vol. 189, no. 3, pp. 1171–1175. ISSN 00219193. Available at: DOI: 10.1128/JB.01589-06

SUNAGAWA, S., COELHO, L. P., CHAFFRON, S., KULTIMA, J. R., LABADIE, K., SALAZAR, G., DJAHANSCHIRI, B., ZELLER, G., MENDE, D. R., ALBERTI, A., CORNEJO-CASTILLO, F. M., COSTEA, P. I., CRUAUD, C., D'OVIDIO, F., ENGELEN, S., FERRERA, I., GASOL, J. M., GUIDI, L., HILDEBRAND, F., KOKOSZKA, F., LEPOIVRE, C., LIMA-MENDEZ, G., POULAIN, J., POULOS, B. T., ROYO-LLONCH, M., SARMENTO, H., VIEIRA-SILVA, S., DIMIER, C., PICHERAL, M., SEARSON, S., KANDELS-LEWIS, S., BOSS, E., FOLLOWS, M., KARP-BOSS, L., KRZIC, U., REYNAUD, E. G., SARDET, C., SIERACKI, M., VELAYOUDON, D., BOWLER, C., DE VARGAS, C., GORSKY, G., GRIMSLEY, N., HINGAMP, P., IUDICONE, D., JAILLON, O., NOT, F., OGATA, H., PESANT, S., SPEICH, S., STEMMANN, L., SULLIVAN, M. B., WEISSENBACH, J., WINCKER, P., KARSENTI, E., RAES, J., ACINAS, S. G., BORK, P. 2015. Structure and function of the global ocean microbiome. In: *Science*. 2015, vol. 348, no. 6237, pp. 1–10. ISSN 10959203. Available at: DOI: 10.1126/science.1261359

TEMPERLEY, R. J., SENECA, S. H., TONSKA, K., BARTNIK, E., BINDOFF, L. A., LIGHTOWLERS, R. N., CHRZANOWSKA-LIGHTOWLERS, Z. M. A. 2003. Investigation of a pathogenic mtDNA microdeletion reveals a translation-dependent deadenylation decay pathway in human mitochondria. In: *Human Molecular Genetics*. 2003, vol. 12, no. 18, pp. 2341–2348. ISSN 09646906. Available at: DOI: 10.1093/hmg/ddg238

THOMAS, P. D., CAMPBELL, M. J., KEJARIWAL, A., MI, H., KARLAK, B.,

DAVERMAN, R., DIEMER, K., MURUGANUJAN, A., NARECHANIA, A. 2003. PANTHER: A library of protein families and subfamilies indexed by function. In: *Genome Research*. 2003, vol. 13, no. 9, pp. 2129–2141. ISSN 10889051. Available at: DOI: 10.1101/gr.772403

THOMSEN, H. A., BUCK, K. R., BOLT, P. A., GARRISON, D. L. 1991. Fine structure and biology of Cryothecomonas gen. nov. (Protista incertae sedis) from the ice biota . In: *Canadian Journal of Zoology*. 1991, vol. 69, no. 4, pp. 1048–1070. ISSN 0008-4301. Available at: DOI: 10.1139/z91-150

THOMSEN, H. A., LARSEN, J. 1993. The ultrastructure of Commation gen. nov. (Stramenopiles incertae sedis), a genus of heterotrophic nanoplanktonic flagellates from antarctic waters. In: *European Journal of Protistology*. 1993, vol. 29, no. 4, pp. 462–477. ISSN 09324739. Available at: DOI: 10.1016/S0932-4739(11)80409-8

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., STRASSERT, J. F. H., KOLÍSKO, M., BROWN, M. W. 2021. PhyloFisher: A phylogenomic package for resolving eukaryotic relationships. In: *PLOS Biology*. 2021, vol. 19, no. 8, pp. 1–22. ISSN 1545-7885. Available at: DOI: 10.1371/journal.pbio.3001365

TIMMIS, J. N., AYLIFF, M. A., HUANG, C. Y., MARTIN, W. 2004. Endosymbiotic gene transfer: Organelle genomes forge eukaryotic chromosomes. In: *Nature Reviews Genetics*. 2004, vol. 5, no. 2, pp. 123–135. ISSN 14710056. Available at: DOI: 10.1038/nrg1271

VALACH, M., GONZALEZ ALCAZAR, J. A., SARRASIN, M., LANG, B. F., GRAY, M. W., BURGER, G. 2021. An Unexpectedly Complex Mitoribosome in Andalucia godoyi, a Protist with the Most Bacteria-like Mitochondrial Genome. In: *Molecular biology and evolution*. 2021, vol. 38, no. 3, pp. 788–804. ISSN 15371719. Available at: DOI: 10.1093/molbev/msaa223

VALACH, M., LÉVEILLÉ-KUNST, A., GRAY, M. W., BURGER, G. 2018. Respiratory chain complex I of unparalleled divergence in diplonemids. In: *Journal of Biological Chemistry*. 2018, vol. 293, no. 41, pp. 16043–16056. ISSN 1083351X. Available at: DOI: 10.1074/jbc.RA118.005326

VOSSEBERG, J., VAN HOOFF, J. J. E., MARCET-HOUBEN, M., VAN VLIMMEREN, A., VAN WIJK, L. M., GABALDÓN, T., SNEL, B. 2021. Timing the origin of eukaryotic cellular complexity with ancient duplications. In: *Nature Ecology and Evolution*. 2021, vol.

5, no. 1, pp. 92–100. ISSN 2397334X. Available at: DOI: 10.1038/s41559-020-01320-z

WALKER, G., DACKS, J. B., MARTIN EMBLEY, T. 2006. Ultrastructural description of Breviata anathema, N. gen., N. Sp., the organism previously studied as "Mastigamoeba invertens." In: *Journal of Eukaryotic Microbiology*. 2006, vol. 53, no. 2, pp. 65–78. ISSN 10665234. Available at: DOI: 10.1111/j.1550-7408.2005.00087.x

YABUKI, A., CHAO, E. E., ISHIDA, K. I., CAVALIER-SMITH, T. 2012. Microheliella maris (Microhelida ord. n.), an Ultrastructurally Highly Distinctive New Axopodial Protist Species and Genus, and the Unity of Phylum Heliozoa. In: *Protist*. 2012, vol. 163, no. 3, pp. 356–388. ISSN 14344610. Available at: DOI: 10.1016/j.protis.2011.10.001

YABUKI, A., EIKREM, W., TAKISHITA, K., PATTERSON, D. J. 2013. Fine Structure of Telonema subtilis Griessmann, 1913: A Flagellate with a Unique Cytoskeletal Structure Among Eukaryotes. In: *Protist*. 2013, vol. 164, no. 4, pp. 556–569. ISSN 16180941. Available at: DOI: 10.1016/j.protis.2013.04.004

YANG, L., DUFF, M. O., GRAVELEY, B. R., CARMICHAEL, G. G., CHEN, L. L. 2011. Genomewide characterization of non-polyadenylated RNAs. In: *Genome Biology*. 2011, vol. 12, no. 2, p. R16. ISSN 14747596. Available at: DOI: 10.1186/gb-2011-12-2-r16

YUBUKI, N., ČEPIČKA, I., LEANDER, B. S. 2016. Evolution of the microtubular cytoskeleton (flagellar apparatus) in parasitic protists. In: *Molecular and Biochemical Parasitology*. 2016, vol. 209, nos. 1–2, pp. 26–34. ISSN 18729428. Available at: DOI: 10.1016/j.molbiopara.2016.02.002

YUBUKI, N., LEANDER, B. S. 2013. Evolution of microtubule organizing centers across the tree of eukaryotes. In: *Plant Journal*. 2013, vol. 75, no. 2, pp. 230–244. ISSN 09607412. Available at: DOI: 10.1111/tpj.12145

YUBUKI, N., ZADROBÍLKOVÁ, E., ČEPIČKA, I. 2017. Ultrastructure and Molecular Phylogeny of Iotanema spirale gen. nov. et sp. nov., a New Lineage of Endobiotic Fornicata with Strikingly Simplified Ultrastructure. In: *Journal of Eukaryotic Microbiology*. 2017, vol. 64, no. 4, pp. 422–433. ISSN 15507408. Available at: DOI: 10.1111/jeu.12376

ZAREMBA-NIEDZWIEDZKA, K., CACERES, E. F., SAW, J. H., BÄCKSTRÖM, Di., JUZOKAITE, L., VANCAESTER, E., SEITZ, K. W., ANANTHARAMAN, K., STARNAWSKI, P., KJELDSEN, K. U., STOTT, M. B., NUNOURA, T., BANFIELD, J. F., SCHRAMM, A., BAKER, B. J., SPANG, A., ETTEMA, T. J. G. 2017. Asgard archaea illuminate the origin of eukaryotic cellular complexity. In: *Nature*. 2017, vol. 541, no. 7637,

pp. 353–358. ISSN 0028-0836. Available at: DOI: 10.1038/nature21031

ZHOU, Y., WANG, Z., ZHANG, J., GU, F. 2011. Ultrastructure of extrusomes in hypotrichous ciliate Pseudourostyla nova. In: *Chinese Journal of Oceanology and Limnology*. 2011, vol. 29, no. 1, pp. 103–108. ISSN 02544059. Available at: DOI: 10.1007/s00343-011-9943-7

ZOLOTAREV, V. A., MYL'NIKOVA, Z. M., MYL'NIKOV, A. P. 2011. The ultrathin structure of amoeboid flagellate Thaumatomastix sp. (Thaumatomonadida (Shirkina) Karpov, 1990). In: *Inland Water Biology*. 2011, vol. 4, no. 3, pp. 287–292. ISSN 19950829. Available at: DOI: 10.1134/S1995082911020222