Genomics of Preaxostyla Flagellates Illuminates Evolutionary Transitions and the Path Towards Mitochondrial Loss

Lukáš V. F. Novák[*1,2], Sebastian C. Treitli[1], Jan Pyrih[1], Paweł Hałakuc[3], Shweta V. Pipaliya[4,7], Vojtěch Vacek[1], Ondřej Brzoň[1], Petr Soukal[1], Laura Eme[2], Joel B. Dacks[4,5], Anna Karnkowska[3], Marek Eliáš[6] & Vladimír Hampl[*1]

[1] Charles University, Faculty of Science, Department of Parasitology, BIOCEV, Vestec, Czech Republic.

[2] Ecology, Systematics, and Evolution Unit, Université Paris-Saclay, CNRS, Orsay, France.

[3] Institute of Evolutionary Biology, Faculty of Biology, Biological and Chemical Research Centre, University of Warsaw, Poland.

[4] Division of Infectious Diseases, Department of Medicine, University of Alberta, Edmonton, Canada.

[5] Institute of Parasitology, Biology Centre, Czech Academy of Sciences, České Budějovice, Czechia.

[6] University of Ostrava, Faculty of Science, Department of Biology and Ecology, Ostrava, Czech Republic.

[7] School of Life Sciences, Ecole Polytechnique Federale de Lausanne, Lausanne, Switzerland; Swiss Institute of Bioinformatics, Lausanne, Switzerland.

21 * Corresponding author

ABSTRACT

The notion that mitochondria cannot be lost was shattered with the report of an oxymonad *Monocercomonoides exilis*, the first eukaryote arguably without any mitochondrion. Yet, questions remain about whether this extends to the entire oxymonad lineage and how this transition took place. The Oxymonadida is a group of gut endobionts taxonomically housed in the Preaxostyla which also contains free-living flagellates of the genera *Trimastix* and *Paratrimastix*. The latter two taxa harbour conspicuous mitochondrion-related organelles (MROs). Here we report high-quality genome and transcriptome assemblies of two Preaxostyla representatives, the free-living *Paratrimastix pyriformis* and the oxymonad *Blattamonas nauphoetae*. We note that *P. pyriformis* possesses a set of unique or ancestral features among metamonads or eukaryotes, e.g., *p*-cresol synthesis, NAD^+ synthesis, selenium volatilization, or mercury methylation, demonstrating the biochemical versatility of this protist lineage. We performed thorough comparisons among all available genomic and transcriptomic data of Preaxostyla to further decipher the evolutionary changes towards amitochondriality, endobiosis and unstacked Golgi. Our results provide insights into the metabolic and endomembrane evolution, but most strikingly the data confirm the complete loss of mitochondria for all three oxymonad species investigated (*M. exilis*, *B. nauphoetae*, and *Streblomastix strix*), suggesting the amitochondriate status may be common to Oxymonadida.

39 **INTRODUCTION**

40 Multiple eukaryotic lineages have adapted to low-oxygen and/or endobiotic lifestyles by modifying
41 their mitochondria into a wide range of mitochondrion-related organelle (MRO) types via the
42 process of reductive and adaptive evolution (Roger et al. 2017). The most radically modified MROs
43 are traditionally categorized as hydrogenosomes, producing ATP by extended glycolysis, and as
44 mitosomes with no role in energy metabolism (Müller et al. 2012). However, exploration of a
45 broader diversity of MRO-containing lineages makes it clear that MROs of various organisms form
46 a functional continuum (Stairs et al. 2015; Klinger et al. 2016; Leger et al. 2017; Onuț-Brännström
47 et al. 2023). The extreme outcome of mitochondrial reductive evolution is the complete loss of the
48 organelle, but so far only one organism has been conclusively shown to have reached this point,
49 the chinchilla gut symbiont *Monocercomonoides exilis* (Oxymonadida, Preaxostyla, Metamonada).
50 A genomic project has thoroughly corroborated the amitochondriate status of *M. exilis*, in which it
51 failed to identify any mitochondrion-associated genes while showing multiple other eukaryotic
52 cellular systems to be well represented (Karnkowska et al. 2016; Karnkowska et al. 2019; Treitli
53 et al. 2021).

54 Oxymonadida contains approximately 140 described species of morphologically divergent and
55 diverse flagellates exclusively inhabiting the digestive tracts of metazoans, of which none has been
56 shown to possess a mitochondrion by cytological investigations (Hampl 2017). Therefore, the
57 entire Oxymonadida group may be an amitochondriate lineage, and the mitochondrion was lost in
58 its common ancestor. Oxymonads belong to Preaxostyla, one of the five major clades of the phylum
59 Metamonada consisting of exclusively anaerobes or microaerophiles; the other clades being
60 represented by Fornicata (e.g., *Giardia intestinalis*), Parabasalia (e.g., *Trichomonas vaginalis*),
61 barthelonids and their relatives (the BRC group; Yazaki et al. 2020; Williams et al. 2023) and

62 Anaeramoebidae (e.g., *Anaeramoeba flamelloides*; Stairs et al. 2021). The two additional known
63 branches of Preaxostyla, classified as the genera *Trimastix* and *Paratrimastix*, split out at two
64 successive points off the trunk leading to Oxymonadida. They are both comprised of free-living,
65 bacterivorous flagellates exhibiting a typical excavate morphology and ultrastructure and thriving
66 in low-oxygen environments (Zhang et al. 2015).

67 *Paratrimastix pyriformis* has been shown to possess an MRO morphologically resembling a
68 hydrogenosome. However, it is likely not involved in the ATP-generating extended glycolysis but
69 plays a role in the one-carbon metabolism of the cell (Hampl et al. 2008; Zubáčová et al. 2013;
70 Zítek et al. 2022). Putative MROs have also been observed in electron microscopy studies of a
71 *Trimastix* representative, *T. marina* (Zhang et al. 2015), and their nature was illuminated by
72 transcriptomic data (Leger et al. 2017). As a group with at least two MRO-bearing lineages and an
73 amitochondriate species/clade nested within, the Preaxostyla clade provides a promising model
74 system to study the causes, conditions, and consequences of the loss of this ubiquitous cellular
75 compartment.

76 To start answering questions about the timing and circumstances of this single evolutionary
77 experiment in which the mitochondrion was lost from the cell, a denser sampling of omics-level
78 data from Preaxostyla is needed. Currently, the available data encompasses a genome draft of
79 *M. exilis*, transcriptomes of *P. pyriformis* and *T. marina* with variable completeness and a
80 fragmentary single-cell genome assembly of an oxymonad *Streblomastix strix*. As a key step
81 towards this end, we present high-quality genomic assemblies for another oxymonad, *Blattamonas*
82 *nauphoetae*, and for the free-living *P. pyriformis*. The former has enabled us to test the hypothesis
83 of amitochondriality as a shared feature of oxymonads in general, while the latter provided deeper
84 insights into the metabolic capabilities of free-living Preaxostyla represented by *P. pyriformis*. The

85 distribution of mitochondrion hallmark proteins and comparisons of gene repertoires and metabolic
86 functions among five Preaxostyla species of various ecology and MRO status were used to
87 illuminate the adaptations connected to amitochondriality and the origin of endobiotic lifestyle
88 within the group.

89 RESULTS AND DISCUSSION

90 The new genome assemblies of *Paratrimastix pyriformis* and *Blattamonas nauphoetae* are 91 highly contiguous and nearly complete

92 We employed a combination of Oxford Nanopore and Illumina technologies to obtain genome
93 assemblies of two species of Preaxostyla. As both species are grown in polyxenic cultures where
94 the eukaryotes represent a minority of cells, we employed multiple rounds of decontamination.
95 Prior to the genomic DNA isolation, enrichment of the sample for the eukaryotic cells was achieved
96 using filtration, and after sequencing the data were carefully decontaminated bioinformatically,
97 resulting in two highly contiguous eukaryotic genome assemblies (see Materials and Methods).
98 The basic characteristics of these genomic assemblies and their comparison to the previously
99 published assemblies of Preaxostyla taxa are given in Table 1.

100 The *B. nauphoetae* genome was assembled into 879 contigs spanning 88,537,989 bp, with an
101 N50 = 199,589 bp and a GC content of 44.96%. Automatic and manual gene prediction resulted in
102 25,221 predicted protein-coding genes. The *P. pyriformis* genome was assembled into 633
103 scaffolds spanning 56,627,582 bp, with an N50 = 276,605 bp and a GC content of 60.94%. Manual
104 and automatic gene prediction resulted in 10,815 predicted protein-coding genes. Using BUSCO
105 v3 (Simão et al. 2015) with the eukaryota_odb9 dataset, the genome completeness was estimated
106 to be 76.3% and 76.6% for *P. pyriformis* and *B. nauphoetae*, respectively (Table 1). It should be

107 noted that, despite their wide usage, BUSCO values are not expected to reach 100% in lineages
108 distant from model eukaryotes simply due to the true absence (or high sequence divergence) of
109 some of the assessed marker genes. For example, various Euglenozoa representatives with highly
110 complete genome sequences, including *Trypanosoma brucei*, have BUSCO completeness
111 estimates in the range of 71-88% (Butenko et al. 2020), and representatives of Metamonada fall
112 within the range of 60-91% (Salas-Leiva et al. 2021). Specifically in the case of oxymonad
113 *M. exilis*, the improvement of the genome assembly using long-read resequencing from 2092
114 scaffolds to 101 contigs led to only a marginal increase of BUSCO value from 75.3 to 77.5 (Treitli
115 et al. 2021).

116 **Comparison of gene repertoires within Preaxostyla and contribution of LGT from**
117 **prokaryotes**

118 The data generated in this study together with the previously published data from *M. exilis*
119 (Karnkowska et al. 2019), *S. strix* (Treitli et al. 2019), and *T. marina* (Leger et al. 2017) were
120 exploited to define the common genetic toolkit of Preaxostyla and specific toolkits of its subgroups.
121 We note that the analyses described below rely on a slightly outdated version of the *M. exilis*
122 genome since they had been initiated before a more complete version (Treitli et al. 2021) became
123 available, and we also acknowledge the limitations of employing the inherently incomplete data
124 from *S. strix* (single-cell genome assembly) and *T. marina* (transcriptome assembly only). We
125 created orthologous groups (OGs) from the protein-coding gene inventories of the five Preaxostyla
126 species and 14 other metamonads, omitting the poorly sampled Anaeramoebidae and BRC clade
127 (Stairs et al. 2021) members (Supplementary file 1), and estimated the core gene set for
128 Metamonada and its clade-specific increments in Preaxostyla and Oxymonadida (Fig. 1 and
129 Supplementary file 2). To account for the incomplete nature of the data sets we use the term "core

130 gene set" in a relaxed sense and included genes present in all main subclades (and not necessarily
131 in all species), e.g. in Fornicata, Parabasalia, and Preaxostyla in the case of the Metamonada core
132 gene set (see Material and Methods for details).

133 Each core set can in principle be composed of three categories of genes: 1) genes inherited by the
134 respective group from its earlier eukaryotic ancestors outside Metamonada, i.e., ancestral
135 eukaryotic genes; 2) genes gained by lateral gene transfer (LGT); and 3) group-specific genes that
136 originated in the respective group, i.e., truly novel genes or extremely diverged descendants of
137 ancient genes, the homology of which was not discerned. To roughly distinguish between these
138 categories within each ancestral set, we have searched for protein homologues (e-value < 10^{-15}) in
139 the NCBI non-redundant protein sequence database and in the EukProt database (Richter et al.
140 2022), manually investigated phylogenies of proteins with prokaryotic hits, and in parallel
141 employed an LGT detection pipeline. We note that we restricted the systematic search for candidate
142 gene transfers from prokaryotic donors into Preaxostyla, as eukaryote-to-eukaryote transfers are
143 generally much more difficult to detect and demonstrate conclusively. With this approach we
144 identified hundreds of putative LGT candidate genes in Preaxostyla (Supplementary file 3), which
145 probably originated from at least 190 LGT events involving various prokaryotic donors and
146 affecting different branches of the Metamonada phylogeny (Fig. 1). Of these, 19, 15, 19, 1, and 3
147 LGT events were mapped to the common ancestor of Metamonada, the Fornicata+Preaxostyla
148 clade, Preaxostyla, the *P. pyriformis*+Oxymonadida clade, and Oxymonadida, respectively, i.e.,
149 contributed to the core gene sets of these clades. The numbers of the ancestral LGTs might be
150 underestimated, as they require sufficient sequence similarity between extant proteins to form a
151 clade in a phylogenetic tree and this condition was often not fulfilled.

152 Of the 1,399 OGs from the Metamonada core gene set, 1,398 have homologues outside
153 Metamonada and just a single orthologous group, q2019831, did not have any significant hit from
154 NCBI nr or EukProt database at e-value < 10^{-15}. Interestingly, this protein identified by
155 InterProScan as a member of Major Facilitator Superfamily of transporters is present only in the
156 data set of the parabasalid *Tritrichomonas foetus*, the oxymonad *S. strix* (Streblo_st30458), and the
157 fornicates *Chilomastix cuspidata* and *Dysnectes brevis*. Although it may represent an innovation
158 arising in the common ancestor of Metamonada, its function is clearly not essential. We identified
159 19 LGT events that contributed to the ancestral gene set of Metamonada and further 15 LGT events
160 specific for the Preaxostyla+Fornicata clade. These include metabolic enzymes used in glycolysis,
161 amino acid metabolism and several functionally unrelated pathways. Some of these cases, e.g.,
162 enzymes involved in the arginine deiminase pathway, have been investigated in detail elsewhere
163 (Liapounova et al. 2006; Novák et al. 2016; Karnkowska et al. 2019), while others are proposed
164 here for the first time (Supplementary file 3).

165 11,854 OGs of all investigated metamonads are present only in Preaxostyla (Figure 1B) and 516
166 of these are widespread in this group and represent a putative Preaxostyla-specific increment to the
167 Metamonada core gene set (Figure 1A). 71 of these OGs seem to be indeed Preaxostyla-specific or
168 they have experienced rapid evolution obscuring their origin, as they have no close eukaryotic
169 BLAST hits (e-value < 10^{-15}) in any other organism. The vast majority of these have vague
170 annotations and their function is simply not known. We identified 19 LGT events that contributed
171 to the core gene set of Preaxostyla (Fig. 1, Supplementary file 2 and 3). A previously documented
172 case of LGT into Preaxostyla concerns the proteins involved in the SUF pathway for FeS cluster
173 synthesis (Vacek et al. 2018), and the respective genes are indeed among the LGT candidates
174 detected by our approach.

175 Putting aside the 71 Preaxostyla-unique OGs and the 19 putative Preaxostyla-specific LGTs, the
176 remaining 425 Preaxostyla-specific increments of the core gene set were likely inherited vertically
177 from the eukaryotic common ancestor but lost or not detected in the other metamonad lineages,
178 and as such they represent an interesting addition to the complexity of the common ancestor of
179 Metamonada revealed thanks to the sampling of Preaxostyla genomes. To provide a single specific
180 example, we focus on the small GTPase Rsg1 (Supplementary file 5). This protein was previously
181 characterized in mammalian cells as a ciliogenesis factor (Agbu et al. 2018; Langousis et al. 2022),
182 but its evolutionary origin and taxonomic occurrence have not been investigated. Having noticed
183 its putative homolog in *P. pyriformis*, we searched publicly available sequence databases to
184 demonstrate that Rsg1 orthologs indeed occur outside Metazoa, exhibiting a patchy distribution
185 across distantly related protist lineages (Supplementary file 5). In Metamonada, Rsg1 was found
186 in *P. pyriformis* and *T. marina*, while no orthologs were identified in oxymonads, fornicates,
187 *Barthelona* sp., parabasalids and *Anaeramoeba* spp., apparently due to multiple independent losses
188 of the Rsg1 gene. Notably, Rsg1 orthologs occur only in eukaryotes that are known or predicted to
189 have the capacity to build cilia (Supplementary file 5), indicating that the cilium-associated role
190 defined for the mammalian Rsg1 is likely a general and ancestral functional attribute of the protein.
191 The occurrence of Rsg1 in the free-living Preaxostyla thus points to specific differences in the
192 functioning of their flagella as compared to other metamonads. Furthermore, considering the
193 unsettled position of Metamonada with respect to the root of the eukaryote phylogeny (Derelle et
194 al. 2015), the identification of Rsg1 in metamonads, along with other phylogenetically diverse
195 eukaryotes, strengthens the case that the protein is an ancestral component of the eukaryotic cell
196 functioning already in the last eukaryote common ancestor (LECA).

197 Of the 9,513 OGs present only among oxymonads (Figure 1C), 8,271 seem to be species-specific
198 clusters or singletons while 1,242 were detected more broadly and represent a putative oxymonad-
199 specific increment to the Preaxostyla core gene set (Fig. 1, Supplementary file 2). Of these, 630
200 OGs seem to be oxymonad novelties or divergent ancestral genes, without close BLAST hits (e-
201 value < 10^{-15}) to any of these sequences. We have no or vague annotation for 580 of them which
202 represent a genetic dark matter supported by transcriptome data but bearing no known function.
203 The remaining 50 received annotations usually from KEGG or InterProScan and comprise a
204 mixture of putative functions noticeably enriched for Fe-S cluster containing proteins and proteins
205 involved in endomembrane transfer. The other 612 OGs have a homolog in eukaryotes outside
206 Metamonada in the NCBI non-redundant protein sequence database or EukProt database (e-value
207 < 10^{-15}). We identified three LGT events that contributed to the ancestral gene set of oxymonads
208 and brought proteins with unclear or mutually unrelated functions. One of them is UTP—glucose-
209 1-phosphate uridylyltransferase putatively involved in glycogenesis and cyst wall synthesis (Fig.
210 1, Supplementary file 3).

211 Based on orthology grouping, most of the genetic innovation is observed on the terminal branches
212 of Preaxostyla rather than in the ancestral nodes. Putting aside the highly incomplete transcriptome
213 of *T. marina* and the incomplete and fragmented genome of *S. strix*, we see that species-specific
214 clusters or singletons vary between 906 in *P. pyriformis* to 2,493 in *B. nauphoetae* (Figure 1C, D).
215 Also, 46 LGT events were mapped to the terminal branches and only five to internal nodes of
216 Preaxostyla, suggesting that the acquisition of genes is an ongoing phenomenon, and it might be
217 adaptive to the particular lifestyles of the species. Selected examples of LGTs providing unique
218 capabilities to *P. pyriformis* are discussed below.

219 ***P. pyriformis* possesses unexpected metabolic capacities**

220 The relatively complete and contiguous genomic and transcriptomic datasets for *P. pyriformis*
221 allowed composing a more complete and more accurate inventory of the *P. pyriformis* genetic
222 toolkit than the previous studies (Hampl et al. 2008; Zubáčová et al. 2013). Given that this species
223 represents the first free-living representative of the Preaxostyla lineage studied at the genome level,
224 it is not surprising that our exploration of the *P. pyriformis* genome data revealed several unusual
225 or unexpected features of its metabolism. Below we present in detail a selection of those that are
226 directly related to the anaerobic lifestyle or the specific ecological niche of *P. pyriformis*.

227 *P. pyriformis* encodes a complete pathway required for the biosynthesis of *p*-cresol from tyrosine
228 (Supplementary file 6), only the second reported eukaryote with such capability. This pathway
229 consists of three steps of the Ehrlich pathway (Hazelwood et al. 2008) converting tyrosine to 4-
230 hydroxyphenyl-acetate (AAT, HPPD, ALDH) and the final step catalyzed by a fusion protein
231 comprised of 4-hydroxyphenylacetate decarboxylase (HPAD) and its activating enzyme (HPAD-
232 AE). The *P. pyriformis* HPAD sequence is closely related to its homolog in the free-living
233 archamoebid *M. balamuthi* (Supplementary file 4 A), the only eukaryote reported so far to be able
234 to produce *p*-cresol (Nývltová et al. 2017). Interestingly, we identified additional eukaryotic
235 homologues of this fusion protein in the free-living fornicates *Kipferlia bialata* and *Trepomonas*
236 sp. This distribution may be explained by LGT from prokaryotes followed by one or more LGTs
237 between the eukaryotic lineages facilitated by a shared anaerobic habitat. Interestingly, the close
238 endobiotic relatives of the taxa endowed for *p*-cresol synthesis (*Entamoeba*, oxymonads, and
239 parasitic diplomonads such as *Giardia intestinalis* or *Spironucleus salmonicida*) lack this protein.
240 Biosynthesis of *p*-cresol by *M. balamuthi* was hypothesized to confer a competitive advantage

241 against prokaryotes inhabiting the same environment because this compound inhibits the growth
242 of certain prokaryotes (Nývltová et al. 2017).

243 The presence of demethylmenaquinone methyltransferase / 2-methoxy-6-polyprenyl-1,4-
244 benzoquinol methylase (UbiE) in *P. pyriformis* (Supplementary file 6) is surprising since the
245 organism lacks any other enzymes involved in the ubiquinone and menaquinone biosynthesis.
246 Interestingly, this enzyme has been shown to have an alternative function, a capability to mediate
247 the synthesis of volatile methylated selenium (Se) compounds, namely dimethyl selenide and
248 dimethyl diselenide, in bacteria and plants (Swearingen et al. 2006; Zhou et al. 2009). Selenium is
249 a micronutrient that becomes toxic in high concentrations. Biological Se volatilization is an
250 important detoxification process that converts inorganic Se compounds into volatile organic
251 compounds. We propose that Se volatilization might be the physiological function of the
252 *P. pyriformis* UbiE. This enzyme is another noteworthy case of prokaryotic LGT within the
253 *P. pyriformis* genome, most likely from a verrucomicrobial source (Supplementary file 4 B).

254 *P. pyriformis* has a complete pathway for biosynthesis of NAD^+ from L-aspartate and
255 dihydroxyacetone phosphate consisting of L-aspartate oxidase (NadB), a nicotinate-nucleotide
256 pyrophosphorylase-quinolinate synthase fusion protein (NadC-NadA), nicotinate-mononucleotide
257 adenylyltransferase (NadD), DNA-binding transcriptional repressor/NMN adenylyltransferase
258 (NadR), and NAD^+ synthase (NadE); at least some of these enzymes are also present in *T. marina*
259 (Supplementary file 6). The presence of this pathway in anaerobes is surprising because the
260 standard NadB reaction involves molecular oxygen (O_2). However, NadB can alternatively use
261 fumarate instead of oxygen (Sakuraba et al. 2002). If this is the case also in *P. pyriformis*, it would
262 explain the presence of fumarate hydratase (FH) in this organism: fumarate produced by FH from
263 malate could serve as an electron acceptor for NadB. Phylogenetic analyses of NadA, NadB, NadC,

264 and NadR show the Preaxostyla sequences branch separately from other eukaryotic sequences and
265 have an affinity to Chloroflexi (NadA, NadC) and Lentisphaerae (NadB), hinting at LGT origin
266 from prokaryotes. In the case of NadB, this LGT is likely shared with *M. balamuthi*, which occupies
267 a similar position inside Lentisphaerae. Only the Preaxostyla NadD and NadE genes branch
268 together with other eukaryotes (Supplementary file 4 C–H).

269 Uniquely among eukaryotes, *P. pyriformis* may possess enzymes for mercury (Hg) methylation.
270 Highly toxic methyl mercury is produced in some Bacteria and Archaea (Podar et al. 2015; Cooper
271 et al. 2020). The exact reason why these microorganisms methylate Hg is unknown, however, it is
272 widely accepted that it is not a detoxification process as these are not less susceptible to Hg
273 (Gilmour et al. 2011). In prokaryotes, Hg methylation is performed by two enzymes HgcA and
274 HgcB encoded by two independent genes, although they are often found adjacent to each other in
275 a single operon (Christensen et al. 2016). In contrast, *P. pyriformis* contains a single fusion protein
276 HgcAB (Supplementary file 6). As the order of the two parts reflects the arrangement of the genes
277 in bacteria, both were likely acquired *en bloc* by lateral transfer of a whole operon (Supplementary
278 file 4 I, J). The structure of the prokaryotic HgcAB complex was recently determined (Cooper et
279 al. 2020). Comparison of PAPYR_7512 to HgcA and HgcB from *Pseudodesulfovibrio mercurii*
280 and a selection of other prokaryotes revealed remarkable similarity and conservation and contains
281 all important residues necessary for methylation of Hg, a function that has not been reported in any
282 other eukaryote (Fig. 2).

283 Altogether, the gene inventory of *P. pyriformis* demonstrates that this protist possesses several
284 metabolic capabilities which are very rare or so far unseen in eukaryotes in general or in other
285 metamonads, and that LGT is a common means of their acquisition. Future explorations of
286 genomes of additional free-living preaxostylans will allow us to assess which of these traits are

287 specific for the *P. pyriformis* lineage and which are instead plesiomorphic in Preaxostyla, being
288 thus part of the functional context in which the evolutionary transition to oxymonads happened.

289 **Membrane transporter complement may reflect adaptation to an endobiotic lifestyle in**
290 **oxymonads**

291 The collected -omic data set of Preaxostyla allows to examine several interesting evolutionary
292 transitions. The first of them is the transition from a free-living to an endobiotic lifestyle. As an
293 assessment of this, we examined proteins responsible for the transport of metabolites and other
294 chemicals across the plasma membrane and other cell membranes. We searched for a broad
295 spectrum of transmembrane transporters (except for transporters of MROs) using homology-based
296 methods in order to compare the repertoire of functional types as well as the number of paralogues
297 between the five species of Preaxostyla (Supplementary file 7).

298 The most gene-rich group of membrane transporters identified in Preaxostyla is the ATP-binding
299 cassette (ABC) superfamily represented by MRP and pATPase families, just like in *T. vaginalis*
300 (Carlton et al. 2007). Altogether, representatives of 19 transporter families have been identified in
301 Preaxostyla. All of them are present in both free-living species *P. pyriformis* and *T. marina*, while
302 four families (PotE, SPNS, RFC, and TauE) are missing in all three endobiotic oxymonads
303 investigated. On the other hand, transporters of nucleosides, sugars, amino acids, choline, and
304 phospholipids have consistently higher numbers of paralogues in the genomes of the oxymonads
305 *M. exilis* and *B. nauphoetae* than in *P. pyriformis* and their phylogenetic trees are consistent with
306 gene family expansion in oxymonads (Supplementary file 4 K–O). These two observations parallel
307 the findings of functional domain loss and expansion by gene duplication of transporter families in
308 Microsporidia (Nakjang et al. 2013), possibly hinting at a broader evolutionary pattern at the
309 transition to the endobiotic lifestyle.

310 **Evidence for the Golgi complex in all Preaxostyla is consistent with the cisternal adhesion**
311 **model of Golgi stacking**

312 Another evolutionary transition addressable by our results is the presence of a morphologically
313 identifiable Golgi body in *P. pyriformis* and *T. marina* versus the lack in oxymonads (O'Kelly et
314 al. 1999; Zhang et al. 2015; Treitli et al. 2018). Nonetheless, a substantial complement of proteins
315 associated with Golgi membrane-trafficking and transport has been previously reported,
316 confirming at least the genomic signal for a Golgi body, in the oxymonad *M. exilis* (Karnkowska
317 et al. 2016; Karnkowska et al. 2019). Therefore, our dataset represents one of the tightest samplings
318 to date where genomic data is available for the closest known species on both sides of the divide
319 between stacked and unstacked Golgi morphology. Consequently, we searched in our datasets to
320 assess the Golgi complement in the additional Preaxostyla representatives, particularly to see
321 whether the complement was more extensive in the organisms possessing stacked Golgi bodies.

322 The scope of our examination included proteins involved in vesicle formation, vesicle fusion, and
323 the golgin proteins implicated in Golgi structure (Boncompain and Weigel 2018; Ahat et al. 2019;
324 Kulkarni-Gosavi et al. 2019; Li et al. 2019; Park et al. 2021; Aridor 2022). We observed near
325 complete complements of the COPI, AP1, AP3, AP4, and Retromer vesicle coats, as well as the
326 GARP complex, Trs120, and syntaxins 5 and 16 (Fig. 3, Supplementary file 6). We also noted at
327 least one golgin protein in each of the organisms. Indeed, we observed additional paralogues of the
328 vesicle trafficking machinery (e.g., AP1, Retromer, GARP/EARP) in oxymonads compared to
329 *P. pyriformis* and *T. marina*, (Fig. 3, Supplementary file 6). These data, together with previously
330 published observations (Vargová et al. 2021), are indicative of Golgi bodies with multiple
331 anterograde and retrograde pathways entering and exiting the organelle present in all Preaxostyla
332 species.

333 We did observe two clear differences in the sets of Golgi-associated proteins between the stacked
334 and unstacked possessing organisms. Firstly, *P. pyriformis* encodes seven of the eight Conserved
335 Oligomeric Golgi (COG) complex proteins, while only a sparse representation of the COG complex
336 was seen in the oxymonads. Secondly, the golgin CASP was found in both *P. pyriformis* and
337 *T. marina*, but in none of the oxymonad genomes. Golgin 84 was also found in *P. pyriformis* alone.
338 This marks the first report of CASP or golgin 84 from a metamonad (Barlow et al. 2018) suggesting
339 independent losses of these proteins in the Oxymonadida, Parabasalia, and Fornicata lineages.
340 While caution is warranted when reporting the absence of any given single protein from any given
341 genome, our observations do show a greater number of encoded Golgi-structure implicated proteins
342 in the stacked-Golgi possessing lineages than in the oxymonads (4, 7 vs 1, 4, 3 respectively; Fig. 3).
343 Though expression levels would need to be taken into account, this observation is nonetheless
344 consistent with the cisternal adhesion model (Lee et al. 2014), i.e., that it is the amount of adhesive
345 golgin-type proteins that regulate stacking rather than the identity of any given Golgi-stacking
346 protein (Barlow et al. 2018).

347 **Absence of mitochondrion and its effect on the inventory of enzymes for extended glycolysis,**
348 **amino acid metabolism and FeS cluster-containing proteins**

349 The uniqueness of Preaxostyla lies in the fact that it includes species harbouring MROs,
350 exemplified by *T. marina* and *P. pyriformis* (Hampl et al. 2008; Zubáčová et al. 2013; Leger et al.
351 2017), and at least one species lacking any form of a mitochondrion, *M. exilis* (Karnkowska et al.
352 2016). In the remaining parts of the manuscript, we inspect how the loss of mitochondria has seeped
353 into the inventory of enzymes for extended glycolysis, amino acid metabolism and those containing
354 FeS clusters, as the function or maturation of these proteins is often connected to mitochondrion or

355 MRO. Finally, we thoroughly test the presence of mitochondrion in other investigated oxymonads
356 to establish whether its absence is unique for *M. exilis*.

357 Many eukaryotic anaerobes, including Preaxostyla, generate ATP using the extended glycolysis
358 pathway, which produces acetate, CO_2, and H_2 from pyruvate while performing substrate-level
359 phosphorylation of ADP to ATP (Lindmark and Müller 1973). The pathway uses pyruvate as a
360 substrate, which is either directly sourced from glycolysis or produced by decarboxylation of
361 malate through the activity of the malic enzyme (ME), which has been identified in all five
362 Preaxostyla species. All Preaxostyla apparently rely on oxidative decarboxylation of pyruvate to
363 acetyl coenzyme A (acetyl-CoA) and CO_2 in a reaction catalyzed by pyruvate:ferredoxin
364 oxidoreductase (PFO; Williams et al. 1987), owing to the identification of three to six PFO
365 isoforms in each species analyzed (Fig. 4, Supplementary file 4 P, Supplementary file 6). None of
366 the alternative enzymes mediating the conversion of pyruvate to acetyl-CoA, pyruvate:NADP+
367 oxidoreductase (PNO) and pyruvate formate lyase (PFL), could be detected in any of the studied
368 species.

369 Both the decarboxylation of malate by ME and of pyruvate by PFO are oxidative processes that
370 release electrons, producing NADH and reduced ferredoxin, respectively, and these electron
371 carriers need to be reoxidized. The final fate of the electrons carried by ferredoxin often lies in the
372 reduction of protons to molecular hydrogen through the activity of [FeFe] hydrogenases (HydA;
373 Payne et al. 1993). In addition to the "simple" hydrogenases, which are present in all species of
374 Preaxostyla, [FeFe] hydrogenases with N-terminal homology to the NuoG subunit of NADH-
375 quinone oxidoreductase and a C-terminal homology to NADPH-dependent sulfite reductase
376 (CysJ), were identified in MRO containing *T. marina* and *P. pyriformis* (Fig. 5). Similar "fused"
377 hydrogenases (Fig. 5) have been previously reported in other eukaryotic anaerobes, including

378 *T. vaginalis* (Tachezy and Doležal 2007), the breviate *Pygsuia biforma* (Stairs et al. 2014), and the
379 jakobid *Stygiella incarcerata* (Leger et al. 2016). Although they do not belong to the group of A3
380 trimeric hydrogenases (Greening et al. 2016) known to be capable of NADH oxidation via electron
381 confurcation (Schut and Adams 2009), they were hypothesized to catalyze NAD(P)H-dependent
382 formation of H_2 (Tachezy and Doležal 2007). Although the presence of these large hydrogenases
383 correlates with the presence of MRO, they were not detected in the MRO proteome of *P. pyriformis*
384 (Zítek et al. 2022) and their localization is unknown.

385 The second reaction of the extended glycolysis, which yields ATP, acetate, and CoA, can be
386 catalyzed either by a two-enzyme system consisting of acetate:succinate CoA-transferase (ASCT)
387 and succinyl CoA synthetase (SCS) like in *T. vaginalis*, or by a single enzyme acetyl-CoA
388 synthetase (ACS) like in *G. intestinalis*. All five Preaxostyla species contain ACS, while neither
389 ASCT nor SCS was identified. ACS has a complicated evolutionary history in Metamonada
390 characterized by multiple LGT events and gene losses (Leger et al. 2017; Yazaki et al. 2020;
391 Vargová et al. 2022). The majority of ACS homologues in Metamonada are predicted to function
392 in the cytosol, with the only exception of the ACS2 isoform from *S. salmonicida* that functions in
393 the MRO (Jerlström-Hultqvist et al. 2013; Leger et al. 2017). Phylogenetic analysis of Preaxostyla
394 ACSs (Supplementary file 4 Q) shows four unrelated clades, none in close relationship to the
395 *S. salmonicida* MRO homolog, consistent with our assumption that these enzymes are cytosolic in
396 Preaxostyla.

397 **Amino acid metabolism is richer in *Trimastix* and *Paratrimastix* than in oxymonads**

398 We noticed clear differences between oxymonads on one side and MRO-containing species
399 *T. marina* and *P. pyriformis* on the other with respect to their amino-acid metabolism. Hypothetical
400 metabolic maps of the amino acid metabolism were reconstructed based on the metabolic maps in

401 the KEGG database and catalytic activities of enzymes reported from other metamonads (Carlton
402 et al. 2007; Xu et al. 2014).

403 Of note is the part of amino acid metabolism connected to the folate and methionine cycle.
404 Methionine (in the form of S-adenosylmethionine, SAM), consumed in reactions catalyzed by
405 SAM-dependent methyltransferases, can be regenerated in the methionine cycle, which is present
406 in *P. pyriformis* and *T. marina* but not identified as encoded in oxymonads. The presence of a
407 methionine cycle is probably connected with the presence of MROs containing a complete glycine
408 cleavage system (GCS) and serine hydroxymethyl transferase (SHMT) in *P. pyriformis* (Zítek et
409 al. 2022) and hypothetically also in *T. marina*. The methyl residues liberated from glycine and
410 serine inside MRO enter the folate cycle and can be later utilized for the methylation of
411 homocysteine to produce methionine via MetH (Ducker and Rabinowitz 2017). The
412 transsulfuration pathway, associated with the folate and methionine cycles in mammals (Stipanuk
413 1986), was not found in any Preaxostyla species. It is likely that the loss of the folate and
414 methionine cycles in oxymonads is tightly connected with the loss of MRO.

415 While investigating the Preaxostyla representatives for the presence of components of the glycine
416 cleavage system (GCS) we encountered altogether five homologues of the L protein (GCS-L): two
417 in *P. pyriformis* and one in *T. marina*, *M. exilis*, and *B. nauphoetae* each. Phylogenetic analyses
418 placed them into two clusters (Supplementary file 4 R). The MRO-localized protein of
419 *P. pyriformis* (PAPYR_5544; Zítek et al. 2022) branched with sequences of *T. marina*, Parabasalia
420 and the fornicate *Carpediemonas membranifera* with moderate statistical support, suggesting a
421 common metamonad origin of this supposedly MRO-localized enzymes. The other *P. pyriformis*
422 sequence (PAPYR_1328) and the two oxymonad sequences branched with high statistical support
423 together with Archamoebae (*Mastigamoeba balamuthi* and *Pelomyxa schiedti*) and another

424 eukaryotic anaerobe, *Breviata anathema* (Breviatea); we refer to this divergent clade as GCS-L2.
425 Localization of this protein in *P. pyriformis* could not be established as it was not measured only
426 in one replicate of the experiment (Zítek et al. 2022). The well-supported relationship between the
427 GCS-L2 sequences from Preaxostyla, Archamoeabae, and Breviatea may be explained by a
428 eukaryote-to-eukaryote LGT, although the direction of the transfer among the three eukaryote
429 lineages is unknown. Interestingly, it has been suggested that the GCS-L2 protein could function
430 outside the context of the glycine cleavage system in *M. balamuthi* (Nývltová et al. 2015), which
431 might explain the presence of GCS-L2 protein in the amitochondriate oxymonads. It should be
432 noted that, besides GCS, GCS-L is known to be part of three other protein complexes, pyruvate
433 dehydrogenase, branched-chain amino acid dehydrogenase, and 2-oxoglutarate dehydrogenase
434 (Spalding and Prigge 2010), none of them being present in Preaxostyla. Outside of these typical
435 roles or under specific conditions, GCS-L was shown to have a moonlighting proteolytic activity
436 (Babady et al. 2007) or a diaphorase activity by which it oxidizes NADH using labile ferric iron
437 (Petrat et al. 2003), nitric oxide (Igamberdiev et al. 2004), or ubiquinone (Xia et al. 2001). It is,
438 therefore, possible that one of these moonlighting activities may represent the primary role of GCS-
439 L2.

440 To summarize the capacity to *de novo* synthesize protein-building amino acids, it seems richer in
441 *P. pyriformis* (cysteine, serine, glycine, threonine, and selenocysteine) than in oxymonads
442 (cysteine, serine, and selenocysteine). The inferred capacity of *T. marina* to synthesize only two
443 amino acids (cysteine and selenocysteine) most likely reflects the incompleteness of the sequence
444 data available (Supplementary figs. 1-3). The selenocysteine biosynthesis pathway present in
445 *P. pyriformis*, *T. marina*, and *M. exilis* is notable, as the capacity to synthesize this amino acid has
446 been reported only in *S. salmonicida* among other metamonads studied so far (Xu et al. 2014).

447 Like many other metamonads (Novák et al. 2016), *M. exilis* has been suggested to utilize arginine
448 for ATP production via the arginine deiminase pathway consisting of three enzymes: arginine
449 deiminase (ADI), carbamate kinase (CK), and ornithine transcarbamylase (OTC). This important
450 metabolic capability has been probably formed in the common Metamonada ancestor by the
451 acquisition of genes for ADI and OTC via LGT (Novák et al. 2016). Here we show the presence
452 of the complete arginine deiminase pathway also in *T. marina* and *B. nauphoetae*, while
453 *P. pyriformis* and *S. strix* lack ADI and CK, respectively. This suggests that ATP production via
454 arginine catabolism is widespread but not omnipresent in Preaxostyla, and it is present in both free-
455 living and endobiotic representatives. Other amino acids can be used in energy metabolism as well:
456 cysteine, serine, and tryptophan can be converted to pyruvate, while methionine can be converted
457 to α-keto-butyrate. Both products can be used by PFO and ACS to generate ATP (Anderson and
458 Loftus 2005).

459 **Oxymonads use almost exclusively 4Fe-4S clusters in their iron-sulfur clusters containing**
460 **proteins**

461 Fe-S clusters are ancient and versatile inorganic cofactors present in virtually all living organisms.
462 In a typical eukaryotic cell, Fe-S clusters are synthesized by a combination of the mitochondrial
463 ISC and the cytosolic CIA pathway, but members of Preaxostyla uniquely combine the CIA
464 pathway with the SUF pathway acquired by LGT from bacteria to perform the same task (Vacek
465 et al. 2018; Braymer et al. 2021). The enzymes were shown to localize in the cytosol (Zítek et al.
466 2022) and the first functional details of the Fe-S cluster biogenesis in this group were described
467 recently (Peña-Diaz et al. 2023). This major pathway shift in the Preaxostyla common ancestor was
468 likely a preadaptation for the loss of the mitochondrion in the lineage leading to *M. exilis*
469 (Karnkowska et al. 2016).

470 To assess how much the change in the Fe-S cluster assembly pathway affected the inventory of Fe-
471 S cluster-containing proteins in Preaxostyla, we compared the previously predicted Fe-S proteins
472 of *M. exilis* (Karnkowska et al. 2019) with a set of Fe-S proteins newly predicted from *in silico*
473 proteomes of *P. pyriformis*, *T. marina*, *B. nauphoetae*, and *S. strix* (Supplementary file 8). The
474 numbers of Fe-S cluster-containing proteins identified in individual species varied from 48 in
475 *T. marina* (most likely an underestimate of the real number, owing to the incompleteness of the
476 data) to 93 in *P. pyriformis* and are thus not decreased in comparison to other heterotrophic
477 eukaryotes (Andreini et al. 2016). Predicted Fe-S proteins fell into 164 distinct OGs. The numbers
478 of these OGs shared among Preaxostyla species are shown in Supplementary Fig. S4.

479 Most of the predicted Fe-S proteins in oxymonads contain 4Fe-4S clusters, with the single
480 exception of xanthine dehydrogenase (XDH), which contains both 2Fe-2S and 4Fe-4S clusters.
481 The free-living *P. pyriformis* and *T. marina* furthermore contain other proteins with 2Fe-2S
482 clusters, such as [FeFe] hydrogenases and 2Fe-2S ferredoxin. As 2Fe-2S clusters are more frequent
483 in mitochondrial proteins, the higher number of 2Fe-2S proteins in *P. pyriformis* compared to the
484 oxymonads may reflect the presence of the MRO in this organism.

485 The most widespread types of Fe-S cluster-containing proteins in Preaxostyla fall into expected
486 functional groups acting in processes of extended glycolysis and electron transfer (PFO, HydA,
487 ferredoxins, and flavodoxin-ferredoxin domains containing proteins), DNA replication and repair,
488 transcription and translation (e.g., DNA and RNA polymerases, Rli1p), Fe-S cluster assembly itself
489 (SufDSU, SufB, Nbp35, and NAR-1), and nucleotide (XDH) and amino acid metabolism (L-serine
490 dehydratase).

491 **A systematic search for proteins associated with the mitochondrion fails to identify**
492 **convincing candidates in oxymonads**

493 The hypothesis on the absence of the organelle in *M. exilis* was postulated based on the absence of
494 mitochondrion-related proteins in its genome and transcriptome and led to the twist in the paradigm
495 of mitochondria as ubiquitous among eukaryotes (Karnkowska et al. 2016). Obviously, the
496 amitochondriate status of *M. exilis* is immediately falsifiable by finding any evidence of a putative
497 organelle in this species. Careful inspection of the predicted proteomes of other oxymonads for the
498 presence of mitochondria is an obvious next step that may further support, or weaken, this
499 hypothesis.

500 We used several different methods to identify candidates for mitochondrial proteins (results for all
501 searches are shown in Supplementary file 9 and summarized in Fig. 6). In all cases, proteins
502 identified in the first step were reciprocally searched against the MitoMiner (Smith et al. 2012;
503 Smith and Robinson 2018) and NCBI databases. We considered proteins, whose functional
504 annotation was confirmed by reciprocal BLAST as candidate mitochondrial proteins, which were
505 further manually verified with extensive searches (see Materials and Methods). Proteins having
506 passed this third step were classified as validated mitochondrial proteins.

507 We searched all five investigated species for homologues of nuclear genome-encoded proteins
508 typically associated with mitochondria or MROs in other eukaryotes. Initially, we used profile
509 Hidden Markov Models (HMMs) to search for components of the mitochondrion protein import
510 and maturation machinery, considered one of the most conserved mitochondrial features (Pfanner
511 et al. 2019). Homology searches resulted in 22 (*T. marina*) up to 63 (*M. exilis*) hits, which were
512 further evaluated by reciprocal searches against the MitoMiner and NCBI databases resulting in 57
513 candidates. Reciprocal BLASTs indicated that all the candidates recovered in oxymonad data sets

514 are very likely to be false positives based on the annotations of their top BLAST hits (mainly
515 vaguely annotated kinases, peptidases and chaperones) (Fig. 6, Supplementary file 9). For
516 *P. pyriformis* and *T. marina*, the situation was different. With the single exception of the Tim23
517 protein (PaPyr_1418) other 12 previously identified components of the *P. pyriformis*
518 mitochondrion protein import and maturation machinery (Zubáčová et al. 2013; Zítek et al. 2022)
519 were found in the predicted proteomes: the β-barrel outer membrane translocases Tom40 and
520 Sam50 and the associated receptor Tom22, the inner membrane translocases Tim17, Tim22 and
521 their associated protein Pam18 and Tim44, the α and β subunit of the mitochondrial processing
522 peptidase (MPP), and the chaperone protein HSP70, Cpn60 and Cpn10. Tim17 and Hsp70, were
523 also identified in the *T. marina* dataset, corroborating previous findings (Leger et al. 2017). The
524 successful identification of protein translocon components in these two species validated our
525 approach.

526 MROs export or import ATP and other metabolites typically using transporters from the
527 mitochondrial carrier family (MCF) or sporadically by the bacterial-type (NTT-like) nucleotide
528 transporters (Tsaousis et al. 2008). We did not identify any homolog of genes encoding proteins
529 from these two families in any of the three oxymonads investigated. In contrast, MCF carriers, but
530 not NTT-like nucleotide transporters, were recovered in the number of four for each *P. pyriformis*
531 and *T. marina* (Supplementary file 6).

532 The sensitive homology-based HMM searches were complemented by an extensive search for
533 putative homologues of known mitochondrial proteins using a pipeline based on the MitoMiner
534 database (Supplementary file 9), which was enriched with identified mitochondrial proteins of
535 diverse anaerobic eukaryotes with MROs (see Materials and Methods). As already shown for
536 *M. exilis*, the specificity of the pipeline in organisms with divergent mitochondria is low

537 (Karnkowska et al. 2016). Many of the selected candidates were identified as proteins that are
538 obviously not mitochondrial (e.g., histones or ribosomal proteins) or belong to very general gene
539 ontology categories (e.g., protein-binding). The search recovered a comparable number of proteins
540 as candidates for functions in a putative mitochondrion in all investigated species, from 636
541 candidates in *S. strix* up to 1,025 candidates in *P. pyriformis*. While no reliable mitochondrial
542 protein was found among the oxymonad candidates, for *P. pyriformis* and *T. marina* we identified
543 a list of proteins previously known to be localized in their MROs as well as some new candidates,
544 such as mitochondrial fission protein 1 (Fis1; Gene.12785::gnl|Trimastix_PCT|4043), which is a
545 tail-anchored (TA) protein mediating mitochondrial fission (van der Bliek et al. 2013). The
546 phylogenetic pattern of dynamin-related proteins corroborates the finding of Karnkowska et al.
547 (2019) showing a specific *Paratrimastix+Trimastix* clade potentially mediating the division of
548 MROs in these species (PAPYR_3413 and Gene.668::gnl|Trimastix_PCT|268, Supplementary
549 file 4 S). In addition, both *T. marina* (Gene.3674::gnl|Trimastix_PCT|1191) and *P. pyriformis*
550 (PAPYR_2826) but none of the oxymonads exhibit an orthologue of the mitochondrial outer
551 membrane-anchored protein MIRO (mitochondrial Rho; Eberhardt et al. 2020) broadly conserved
552 in eukaryotes but rare in MRO-containing taxa (Vlahou et al. 2011; Gentekaki et al. 2017); recently
553 it was confirmed by proteomics to be MRO-associated in the latter species (Zítek et al. 2022).

554 As an alternative to homology searches, we also searched for several types of signature sequences
555 typical for mitochondrion-targeted proteins. The matrix proteins of mitochondria and MROs are
556 expected to contain characteristic N-terminal targeting signals (NTS) needed for the targeted
557 import into mitochondrion-related organelles (Dolezal et al. 2006). As we have previously shown
558 that even in the amitochondriate *M. exilis*, a fraction of proteins (though less than 1%) contains
559 NTS recognized by the subcellular localization predicting tools (Karnkowska et al. 2016) and also

560 that 14 *M. exilis* protein can be successfully targeted to the hydrogenosomes of *T. vaginalis*. It has
561 been previously recognized that the presence of a predicted NTS by itself does not prove the
562 targeting, as such amino acid sequences can also appear by chance (Lucattini et al. 2004). Here we
563 used up-to-date tools for targeting signal prediction and identified from two hits in *S. strix* up to
564 248 hits in *B. nauphoteae* (Supplementary file 9). Oxymonad species contain a lower proportion
565 of proteins with the putative mitochondrial NTS (always below 1%) than *Trimastix* and
566 *Paratrimastix* and most of them gave no hits in MitoMiner and the NCBI nr database; the few
567 candidates with good hits matched to proteins unlikely to function in mitochondria. If the NTS in
568 the protein is functional and not formed by chance, it is expected that the NTS will be detected also
569 in its orthologues from other species. With this assumption, we focused on candidates of which
570 more than a single Preaxostyla orthologue was predicted to contain NTS. Only four proteins from
571 oxymonads (ribosomal protein L21, L23a, and L34e, and tRNA-dihydrouridine synthase) fulfilled
572 this criterion, none of them representing a reasonable mitochondrial protein in organisms that
573 certainly lacks a mitochondrial genome. Based on those results, we assume that all NTS predictions
574 on oxymonad proteins are false positives. In *P. pyriformis* and *T. marina* the proportion of proteins
575 with predicted NTS is higher (1.5% and 2.7% respectively), which might reflect the presence of
576 the organelle. Indeed, we identified among them, for example, aminoadipate-semialdehyde
577 dehydrogenase and the GCS-H protein in *T. marina*, both previously suggested to localize to the
578 MRO (Leger et al. 2017).

579 The outer mitochondrial membrane accommodates two special classes of proteins, TA proteins and
580 mitochondrial β-barrel membrane proteins (MBOMPs), the former using specific C-terminal
581 signals (Borgese et al. 2007; Denic 2012; Rada et al. 2019). We identified up to 475 TA hits in the
582 predicted proteome of *S. strix*, and around 100 for the rest of the species, with only 25 for *T. marina*

583 (Supplementary file 9). Depending on the species only five to twelve were considered as candidates
584 with homologues in the MitoMiner database and the majority of these were Golgi apparatus/ER-
585 related. A subset of the TA proteins corresponded to 22 OGs that were present in an oxymonad and
586 at least one more species of Preaxostyla. These were mostly without MitoMiner hits and in most
587 cases annotated as SNARE or other proteins involved in vesicle trafficking. This suggests that the
588 TA prediction did produce true positives, but they all likely represent non-mitochondrial TA
589 proteins. The only probable mitochondrial candidate was Fis1 identified in the predicted proteome
590 of *T. marina* (see above).

591 Our search for proteins with MBOMP characteristics revealed six and 23 hits in *T. marina* and
592 *P. pyriformis*, respectively, but only three of them have a hit in the MitoMiner database, including
593 the *P. pyriformis* Tom40, the detection of which validates our approach. Among the 26 to 36
594 candidate proteins from each oxymonad genome, only four hit a protein in the MitoMiner database,
595 none of the hits being a known MBOMP (Supplementary file 9). Using the same logic as with the
596 previous categories of proteins, only four OGs predicted as MBOMPs were present in an oxymonad
597 and at least one other Preaxostyla species, indicating that the candidates were selected from the
598 species pools nonspecifically and likely do not contain a putative new MBOMP. The only
599 overlapping OGs represent clathrin heavy chain (two OGs), kelch-type beta-propeller, and EF-
600 hand domain-containing protein, which are again rather false positives than mitochondrial proteins.

601 Finally, given the limitations of the previous searches, namely, high false-positive rates, uncertainty
602 about mitochondrial localization of proteins included in the MitoMiner database, and the inability
603 to distinguish between the cytosolic and mitochondrial isoforms, we have used also a more curated
604 approach and searched for orthologues of proteins from the experimentally well-established
605 mitochondrial proteome of *T. brucei* (Panigrahi et al. 2009; Dean et al. 2017; Peikert et al. 2017).

606 The selection of this organism as a reference also considered the fact that it is a member of the
607 eukaryotic supergroup Discoba, which frequently forms a sister group to Metamonada in
608 phylogenomic trees (Tice et al. 2021), making *T. brucei* potentially the closest relative of
609 metamonads out of all eukaryotes with extensively studied mitochondria. Reciprocal BLAST
610 searches of the *T. brucei* mitochondrial proteins against the predicted proteins of Preaxostyla
611 revealed ~200 putative groups of orthologues and those were investigated manually. Careful
612 inspection of the raw localization data in *T. brucei* and protein phylogenies (Supplementary file 10)
613 rejected most cases by disputing either the mitochondrial localization in *T. brucei* or the orthology.
614 The gene orthology was considered uncertain if *T. brucei* and Preaxostyla proteins were separated
615 by prokaryotic or non-mitochondrial eukaryotic homologues in the phylogenetic trees. The 30
616 cases surviving the manual scrutiny (Supplementary file 9) were divided into two groups.

617 The high-confidence group comprises 17 proteins that likely share mitochondrial origin as they
618 form a monophyletic cluster with other mitochondrial homologues. These were present only in
619 *P. pyriformis* and/or *T. marina* and never in oxymonads. Twelve out of these 17 proteins were
620 already predicted to be mitochondrial previously (Zítek et al. 2022). The remaining proteins, such
621 as L-threonine 3-dehydrogenase (TDH) or glycine acetyltransferase thus have a potential for being
622 novel mitochondrial proteins in those two protists.

623 The 13 putative mitochondrial proteins of the low confidence group cluster with mitochondrial
624 proteins in other eukaryotes; however, the mitochondrial origin is not well supported. They either
625 cluster also with cytosolic and peroxisomal isoforms, or the enzymes are known to be dually
626 localized. An example of the former is isocitrate dehydrogenase, for which the yeast cytosolic,
627 peroxisomal, and mitochondrial proteins form a clade in the tree. Examples of the latter are
628 aconitase and alanine aminotransferase, for which a dual cytosolic/mitochondrial location was

629 reported for various eukaryotes including trypanosomatids (Duschak and Cazzulo 1991; Saas et al.
630 2000). This low confidence group contains nine proteins present also in oxymonads
631 (Supplementary file 9), and these are discussed in more detail below.

632 Aspartate and alanine aminotransferases and the two tRNA synthetases in this category are all
633 dually localized in the mitochondrion and cytosol of *T. brucei*. In the case of tRNA synthetases,
634 the dual localization arose specifically in trypanosomatids (Peikert et al. 2017) and, consistently
635 with this hypothesis, Preaxostyla and *T. brucei* proteins cluster together with cytosolic tRNA
636 synthetases of other eukaryotes. In the case of aspartate aminotransferase, the dual localization
637 seems to be of deeper evolutionary origin (Yagi et al. 1990). The two mevalonate pathway enzymes
638 (hydroxymethylglutaryl-CoA synthase and 3-hydroxy-3-methylglutaryl-CoA reductase) are
639 localized in the mitochondrion of kinetoplastids (Peña-Diaz et al. 2004), however, this is again a
640 speciality of this lineage, as eukaryotes typically run this pathway in the cytosol and the ER. In all
641 these cases, we conservatively assume that these enzymes were localized in the cytosol in the
642 common ancestor of Preaxostyla and *T. brucei* and were partially or fully moved to the
643 mitochondrion in the lineage leading to *T. brucei*. The last candidate is the malic enzyme. The
644 evolution of this enzyme family is complicated, and there are two homologues of the enzyme in
645 kinetoplastids, one cytoplasmic and one mitochondrial, which cluster together (Supplementary
646 file 4 T). Again, it cannot be inferred what was the situation in the common ancestor with
647 Preaxostyla.

648 In summary, our systematic searches for MRO proteins allowed us to update the predicted set of
649 MRO-localized proteins of *P. pyriformis* and *T. marina* and to provide updated predictions of their
650 MRO proteomes (Supplementary file 6). Notably, no reliable candidate for an MRO protein was

651 detected in any of the oxymonad data sets, supporting the hypothesis of the absence of the
652 mitochondrion in oxymonads as a group (Fig. 6).

653 **Conclusions**

654 In this manuscript we report the genome sequences of *P. pyriformis* and *B. nauphoetae*, adding to
655 the previously published genomes of *M. exilis* and *S. strix*, thus doubling the number of available
656 annotated genomes of Preaxostyla protists. By comparing the proteomes predicted from these data
657 to available genomes and transcriptomes of metamonads we were able to define the core gene
658 inventories of Metamonada, Preaxostyla, and Oxymonadida, and to detect LGT contributions to
659 them. Based on its genetic toolkit, *P. pyriformis* is endowed with several functions unique among,
660 or rare in, eukaryotes – synthesis of *p*-cresol, mercury methylation, and Se volatilization, or
661 preforms some of the common functions in an unusual way, exemplified by the predicted fumarate-
662 dependent NAD^+ synthesis. We identified differences in inventories of membrane transporters
663 between free-living *P. pyriformis* and *T. marina* and oxymonads that potentially reflect the
664 transition from the free-living to the endobiotic lifestyle at the origin of oxymonads. Proteins
665 involved in the formation and regulation of the Golgi structure have a patchy distribution and show
666 a trend towards loss in oxymonads, which is consistent with the lack of ultrastructural evidence for
667 the presence of a stacked Golgi in oxymonads, but also with an emerging cell biological model for
668 how Golgi maintain their hallmark morphology of stacked cisternae (Lee et al. 2014). The switch
669 to the SUF pathway in these species has apparently not affected the number of Fe-S-containing
670 proteins but led to a decrease in the usage of 2Fe-2S clusters. The loss of MRO impacted
671 particularly the pathways of amino acid metabolism and might relate also to the loss of large
672 hydrogenases in oxymonads.

673 The findings of multiple exclusively mitochondrion-associated proteins in the two studied free-
674 living species, *T. marina* and *P. pyriformis*, corroborate the presence of unique types of MROs in
675 these organisms. Contrarily, the thorough search for mitochondrion-associated proteins has failed
676 to uncover any convincing candidates in the studied oxymonads, further supporting the hypothesis
677 that not only *M. exilis* but at least a large part of oxymonads has completely lost the mitochondrion.
678 This fact moves this unique loss to at least 100 MYA deep past, when oxymonads had been already
679 diversified (Poinar 2009), and shows that a eukaryotic lineage without mitochondria can thrive for
680 eons and undergo pronounced morphological evolution, as is apparent from the range of shapes
681 and specialized cellular structures exhibited by extant oxymonads (Hampl 2017).

682 **MATERIALS AND METHODS**

683 **Cell culture, DNA and RNA isolation**

684 Monoeukaryotic, xenic culture of *P. pyriformis* (strain RCP-MX, ATCC 50935) was maintained
685 in the Sonneborn's Paramecium medium ATCC 802 at room temperature by serial transfer of 1 ml
686 of well-grown culture (approximately 5×10^4 cells/ml) into a 15 ml test tube containing 10 ml of
687 fresh, bacterized medium every week. The medium was bacterized 24 hours before the transfer.
688 *B. nauphoetae* strain NAU3 (Treitli et al. 2018) was cultured in a similar way as described above
689 but using a modified TYSGM media (Diamond 1982).

690 For DNA isolation, 15 liters of *P. pyriformis* and 32 liters of *B. nauphoetae* culture were used. To
691 remove most of the bacterial contamination, the cells were filtered as described previously
692 (Karnkowska et al. 2016). After filtration, the cells were collected at 1200×g for 10 min at 4°C.
693 The DNA was isolated using two different kits. The gDNA samples for PacBio, Illumina HiSeq,
694 and Illumina MiSeq sequencing were isolated using the Qiagen DNeasy Blood & Tissue Kit

695 (Qiagen). The isolated gDNA was further ethanol-precipitated to increase the concentration and
696 remove any contaminants. For Nanopore sequencing DNA was isolated using Qiagen MagAttract
697 HMW DNA Kit (Qiagen) according to the manufacturer's protocol.

698 RNA for Illumina transcriptome sequencing of *P. pyriformis* was isolated from 10 liters of a well-
699 grown culture using TRI reagent (Sigma-Aldrich) The eukaryotic mRNA was selected using the
700 Dynabeads mRNA Purification Kit for mRNA Purification from Total RNA preps (Thermo Fisher
701 Scientific). cDNA was synthesized by the SMARTer PCR cDNA Synthesis Kit (Takara Bio Group)
702 and sequenced using the Illumina HiSeq 2000 sequencer at the Beijing Genomics Institute (BGI).

703 For RNA isolation of *B. nauphoetae* we used 500 ml of a well-grown culture. Prior to RNA
704 isolation, the culture was filtered as described previously (Karnkowska et al. 2016), and the cells
705 were harvested by centrifugation at 1200×g for 10 min at 4ºC. The total RNA was isolated using
706 TRI reagent (Sigma-Aldrich) according to the manufacturer's protocol. From the total RNA,
707 mRNA was isolated using two rounds of Dynabeads Oligo(dT) beads (ThermoFisher) according
708 to the manufacturer's protocol. The purified mRNA was used for *de novo* whole transcriptome
709 sequencing performed at the Beijing Genomics Institute (BGI).

710 **Library preparation, Illumina, and Nanopore sequencing**

711 For *P. pyriformis* PacBio sequencing was performed at SEQme s.r.o. (Dobříš, Czech Republic)
712 using a PacBio RSII sequencer, whereas Illumina sequencing was performed using Illumina HiSeq
713 and MiSeq sequencers at the Institute of Molecular Genetics of the ASCR, v. v. i. (Prague, Czech
714 Republic). For *B. nauphoetae*, one pair-end and two mate-pair libraries were prepared and
715 sequenced on Illumina MIseq PE 2x300bp at Beijing Genomics Institute (BGI; Shenzhen, People's
716 Republic of China).

717 Libraries for Nanopore sequencing were prepared from 4 µg of gDNA for each library. DNA was
718 sheared at ~20kbp using Covaris g-TUBES (Covaris Ltd, UK) and the library was prepared using
719 the ligation sequencing kit from Oxford Nanopore Technologies (SQK-LSK108) according to the
720 manufacturer's protocol. The prepared libraries were loaded onto a R9.4 Spot-On Flow cell (FLO-
721 MIN106) and sequencing was performed on a MinION Mk1B machine for 48 hours using
722 MinKNOW 2.0 software with live base calling enabled. In total we prepared four libraries and used
723 four flow-cells for sequencing, two for *P. pyriformis,* and two for *B. nauphoetae.*

724 **Genome assembly binning and decontamination**

725 The quality of sequencing data was assessed with FastQC (Andrew 2010). For the Illumina data,
726 adapter and quality trimming were performed using Trimmomatic 0.36 (Bolger et al. 2014), with
727 a quality threshold of 15. For the Nanopore data, trimming and removal of chimeric reads were
728 performed using Porechop v0.2.3 (https://github.com/rrwick/Porechop).

729 The initial assembly of the genomes was made only with the Nanopore and PacBio reads using
730 Canu v1.7.1 assembler (Koren et al. 2017), with corMinCoverage and corOutCoverage set to 0 and
731 100000 respectively. After assembly, the data was binned using tetraESOM (Haddad et al. 2009).
732 The resulting eukaryotic bins were also checked using a combination of BLASTn and a scoring
733 strategy based on the identity and coverage of the scaffold as described in Treitli et al. (2019). After
734 binning, the resulting genomic bins were polished in two phases. In the first phase, the scaffolds
735 were polished with Nanopolish (Loman et al. 2015) using the raw reads generated by Nanopore.
736 In the second phase, the resulting scaffolds generated by Nanopolish were further corrected using
737 Illumina short reads with Pilon v1.21 (Walker et al. 2014). Finally, the genome assembly of
738 *P. pyriformis* was further scaffolded with raw RNA-seq reads using Rascaf (Song et al. 2016).

739 **Repeat masking, gene prediction, and annotation of the genomes**

740 Repetitive elements in the genomic assembly were identified using RepeatModeler v1.0.11
741 (Tarailo-Graovac and Chen 2009), and only repeats that are members of known repeat families
742 were used for masking the genome assemblies prior to gene prediction. For the *P. pyriformis* and
743 *B. nauphoetae*, we used Augustus 3.2.3 for gene prediction (Stanke and Waack 2003). For *de novo*
744 prediction of genes, Augustus was first re-trained using a set of gene models manually curated by
745 considering information from mapped transcriptomic sequences and sequence conservation with
746 homologous protein-coding genes. In the next step, intron hints were generated from the RNAseq
747 data and gene prediction was performed on repeat-masked genomes using Augustus. Next,
748 transcriptome assemblies were mapped to the genomes using PASA (Haas et al. 2003) and the
749 resulting assembled transcripts were used as evidence for gene model polishing with EVM (Haas
750 et al. 2008). The genome completeness for each genome was estimated using BUSCO v3 with the
751 Eukaryota odb9 dataset and the genome completeness was estimated on the sets of EVM-polished
752 protein sequences as the input. Details of the BUSCO analysis are given in Supplementary file 11.

753 Automatic functional annotation of the predicted genes was performed using the KEGG Automatic
754 Annotation Server (Moriya et al. 2007), in parallel to similarity searches against NCBI nr protein
755 database using BLASTp. If both gave functional annotation, BLAST was preferred if the e-value
756 was lower than 10^{-30}. Otherwise KEGG was used. Targeted analyses of genes and gene families of
757 specific interest were performed by manual searches of the predicted proteomes using BLASTp
758 and HMMER (Eddy 2011), and complemented by tBLASTn searches of the genome and
759 transcriptome assemblies to check for the presence of individual genes of interest that were
760 potentially missed in the predicted protein sets (single digits of cases per set). Gene models were
761 manually refined for genes of interest when necessary and possible. The annotated genomes were

762 deposited in NCBI under accession number JAPMOS000000000 for *P. pyriformis* and
763 JARBJD000000000 for *B. nauphoetae*.

764 **Defining groups-specific gene sets**

765 We combined the gene inventories of five Preaxostyla species (the two genomes presented here,
766 the *M. exilis* genome under NCBI BioProject PRJNA304271, the *S. strix* genome under NCBI
767 BioProject PRJNA524138, and the *T. marina* transcriptome-derived predicted proteome in
768 EukProt EP00771, Richter et al. 2022) with 14 other metamonads for which the genome or
769 transcriptome drafts were available when the study was initiated (Supplementary file 1, Sheet 1).
770 To optimize the inflation parameter value of OrthoMCL, we performed multiple clustering runs
771 with different inflation parameter values in the range 1.01 to 30.0 and calculated the number of
772 OGs containing genes from all Preaxostyla but no other taxa and from all oxymonads but no other
773 taxa. We chose to proceed with the inflation parameter 4, because under this setting the sum of
774 these two numbers was the highest, so the strength of the clustering should be optimal for our
775 purpose. Under this setting, the complete set of proteins from the 19 species (337,300 items) was
776 assigned to 37,923 groups of genes or singletons (OGs) (Supplementary file 1, Sheet 2). Proteins
777 belonging to these OGs were automatically annotated using BLASTp searches against the NCBI
778 nr protein database to acquire functional annotation (Supplementary file 1, Sheet 3). Venn diagrams
779 were generated using InteractiVenn (Heberle et al. 2015).

780 A relaxed core gene set for Metamonada and its Preaxostyla- and Oxymonadida-specific
781 increments were estimated based on the presence/absence of each OG in the analyzed species
782 (Supplementary file 2). 1,399 OGs present in at least one member of each of the three main
783 metamonad clades represented in the analysis (Parabasalia, Fornicata, and Preaxostyla) were
784 considered as the relaxed core Metamonada gene set. The Preaxostyla specific increment was

785 determined as a union of two sets determined by two different strategies to minimize the effect of
786 data incompleteness: i) the set of 285 OGs present in representatives of all three lineages of
787 Preaxostyla (at least one oxymonad, *P. pyriformis*, and *T. marina*), and ii) the set of 460 OGs
788 present in any four out of the five Preaxostyla species in the analysis. Out of these, 169 OGs were
789 present in all five Preaxostyla species. The Oxymonadida-specific increment was determined as a
790 set of 1,242 OGs present in at least two out of the three oxymonad species yet lacking representative
791 genes in other metamonads analyzed.

792 **Detection of lateral gene transfer candidates**

793 To identify candidate LGT events from prokaryotic donors into Preaxostyla species or any of their
794 metamonad ancestors, we used a phylogenetic approach based on the identification of protein
795 phylogenies in which Preaxostyla sequences (and their other metamonad homologues if they exist)
796 appear nested within a clade of prokaryotic counterparts with strong statistical support. We aimed
797 to be strict in our criteria, preferring underdetection to overprediction of LGT cases.

798 We first discarded predicted Preaxostyla proteins of size smaller than 80 amino acids, since the
799 corresponding phylogenies are unlikely to contain a strong and reliable signal. Given the large
800 number of remaining proteins (109,359), and in order to reduce computational burden, we carried
801 out a preliminary screen to discard sequences without detectable prokaryotic homologues. For this,
802 each preaxostyla sequence was used as a query for a similarity search using DIAMOND (Buchfink
803 et al. 2014) against the GTDB database (release 95; Parks et al. 2022) spanning 194,600 proteomes
804 representing known prokaryotic diversity as of July 2020 (e-value cut-off = 1e-10; identity
805 percentage >25%; query coverage >50%; subject coverage >50%). The query and subject coverage
806 criteria were chosen to ensure we did not consider spurious hit over a short part of the protein.
807 Metamonad sequences with significant hits against the GTDB were then used for a more exhaustive

808 homolog search by using them as queries for DIAMOND searches against the NCBI nr database
809 (same cut-offs as before; 10,000 hits were retained) in order to retrieve both prokaryotic and
810 eukaryotic homologues. To reduce computational burden, we then reduced the number of
811 sequences retrieved from NCBI nr by running CD-HIT (-c 0.7; Li and Godzik 2006) to keep a
812 single representative sequence for clusters of sequences displaying >70% identity. We then
813 retrieved the NCBI taxonomy associated with each sequence.

814 Protein superfamilies were assembled by first running DIAMOND searches of all metamonad
815 sequences against all (-e 1e-20 --id 25 --query-cover 50 --subject-cover 50). Reciprocal hits were
816 gathered into a single FASTA file, as well as their NCBI nr homologues. We also added all
817 metamonad sequences belonging to the same OGs into a single FASTA file.

818 Files with 30 sequences or fewer were discarded, as the chance directionality of the transfer can be
819 determined with any confidence is low when the gene family is represented by a small number of
820 representatives. In addition, files with more eukaryotic homologues than prokaryotic ones were
821 also discarded from follow-up analyses, as they were unlikely to represent a clear case of
822 prokaryote-to-eukaryote gene transfer. Sets of sequences in the remaining files were individually
823 aligned using MAFFT v7.221 (Katoh and Standley 2013), the alignments were trimmed using
824 BMGE (Criscuolo and Gribaldo 2010), and the corresponding phylogenies were inferred using
825 FastTree v.2 (Price et al. 2010).

826 To investigate the potential cases of prokaryotes-to-metamonads LGT, we analysed the unrooted
827 gene trees with the use of the R package PhySortR (Stephens et al. 2016) to identify well-supported
828 clans (potentially monophyletic tree partitions) comprised of metamonad sequences nested within
829 prokaryotic ones. This non-trivial task was done in a stepwise manner by first identifying clans
830 with a statistical support (Shimodaira-Hasegawa test) >0.8 and containing at least 90% of

831 prokaryotes and metamonads. The reason for the latter threshold is to allow for cases of taxonomic
832 misannotation, or for the presence of a few homologues from other eukaryotes, themselves
833 potentially resulting from LGT. After this first assessment, we checked which metamonad
834 sequences formed a clan separated from the rest of the tree by a well-supported bipartition (i.e., a
835 clan with >90% of metamonads with statistical support >0.8) in order to infer if they likely
836 corresponded to a single LGT even within their last common ancestor, or if they were dispersed in
837 the tree suggesting multiple independent transfers of homologous genes.

838 However, when metamonad sequences were detected as paraphyletic (regardless of the possible
839 position of the root), and in order to avoid overestimating the number of LGT events, we manually
840 inspected all phylogenies in order to distinguish most accurately between probable independent
841 cases of LGTs and paraphyly resulting from an overall lack of phylogenetic resolution. For this,
842 our manual inspection took into account: 1) the overall number of sequences in the tree (i.e., a
843 small number can be indicative that only a few homologues passed our DIAMOND threshold
844 because this protein family is fast evolving and that, consequently, the corresponding phylogeny
845 might be poorly resolved and/or not contain representatives of the true diversity of homologues);
846 2) the branch length separating the metamonad sequences (i.e., short branch lengths can be
847 suggestive of an unreliably resolved backbone); 3) the overall taxonomic coherence of the tree (i.e.,
848 do the various clades display taxonomic homogeneity or are the lineages interspersed all over the
849 tree, suggestive of a poorly resolved phylogeny?); 4) the taxonomic affiliation of the prokaryotic
850 neighbors of each independent metamonad clan in the tree (i.e., if they are similar, this is more in
851 favor of a single LGT event ancestral to all metamonad represented in the tree, which phylogenetic
852 reconstructions have not resolved properly); 5) the overall branch support across the tree (i.e., some
853 trees might show a few well supported branches suggestive of paraphyletic metamonads but if the

854 overall backbone of tree is generally poorly supported, we were cautious about the weakness of the
855 phylogenetic signal). Taking all this into account, in case of doubt, we always minimized the
856 number of inferred LGT events by assuming a single event in the last common ancestor of the
857 metamonad sequences in a tree. All phylogenies can be found in Supplementary file 12 (ultrafast
858 Bootstrap support values are indicated at branches. Teal tips represent bacteria, purple are archaea,
859 and coral tips are eukaryotes. Preaxostyla are displayed with a red font).

860 **Single gene phylogenies**

861 The phylogenies of the genes of interest (Supplementary file 4) were analyzed individually using
862 the methodology described below. Eukaryotic homologues of the Preaxostyla genes were gathered
863 by taxonomically constrained iterative BLAST search against publicly available sequence
864 databases in order to sample as broad eukaryotic diversity as possible. In the cases of
865 overrepresented taxa of low interest (e.g. Metazoa, land plants), only a small number of
866 representative sequences were selected arbitrarily. In order to detect potential LGT from
867 prokaryotes while keeping the number of included sequences manageable, prokaryotic homologues
868 were gathered by a BLASTp search with each eukaryotic sequence against the NCBI nr database
869 with an e-value cutoff of 10^{-10} and max. 10 target sequences. The sequences were aligned using
870 MAFFT and automatically trimmed using trimAl v1.2 (Capella-Gutiérrez et al. 2009).
871 Phylogenetic analyses were performed simultaneously using IQ-TREE v2.0.5 (Nguyen et al. 2015)
872 and RAxML-HPC2 on v8.2.12 (Stamatakis 2014) with the LG4X model on the Cyberinfrastructure
873 for the Phylogenetic Research (CIPRES) Science Gateway (Miller et al. 2010). Substitution models
874 were inferred using IQ-TREE TESTNEW function for IQ-TREE.

875 **Analysis of the Rsg1 gene**

876 Candidate Rsg1 orthologues were identified in a variety of databases and other sequence resources
877 (Supplementary file 5) using a combination of BLAST (BLASTp or tBLASTn, depending on the
878 target database) and HMMER searches. The taxon selection essentially followed the list of taxa
879 employed in a recent study of the distribution of the cilium-associated Ras superfamily GTPase
880 Arl16 (Dewees et al. 2022), with further expansion to include additional metamonads and recently
881 sequenced species of isolated phylogenetic position. Additional taxa were added based on
882 searching the newly released EukProt database (Richter et al. 2022) such that representatives of
883 broader taxa consistently possessing or lacking a discernible Rsg1 orthologue were selected and
884 including the final list. Significant BLAST hits retrieved from searches with reference queries (the
885 human sequence – GenBank accession number NP_112169.1, or another already validated
886 sequence from a given taxon to increase the sensitivity of the search in a taxon-specific manner)
887 were evaluated by BLAST searches against a large custom manually annotated database of Ras
888 superfamily genes to distinguish Rsg1 orthologues from other GTPase types. The discrimination
889 was facilitated by Rsg1 having low sequence similarity to other Ras superfamily members (such
890 as Rab GTPases). A set of confidently identified Rsg1 sequences was then used for building a
891 custom profile HMM, which was employed in searches with HMMER to identify possible
892 divergent Rsg1 orthologues missed by BLAST searches. The candidates were again evaluated by
893 BLAST-based comparisons against the custom GTPase database.

894 **Phylogenetic analysis of Golgi-related proteins**

895 Comparative genomics was carried out using HMM searches. Pan-eukaryotic query sequences
896 analyzed and curated for previous pan-eukaryotic vesicle coat, multisubunit tethering complexes,
897 and golgins were used to build profile HMMs. Query sequences were obtained from the

898 supplementary material of previous papers dealing specifically with Adaptins and COPI (Hirst et
899 al. 2014), multisubunit tethering complexes (Klinger et al. 2013), and golgins (Barlow et al. 2018).
900 Individual components and proteins from each sub-complex were aligned using MUSCLE v.3.8.31
901 and the resulting alignment files were used to generate HMM matrices using the HMMBUILD
902 option available through the HMMER package. HMMER searches were carried out in the predicted
903 proteomes of *P. pyriformis*, *B. nauphoetae*, and *S. strix*, whereas for *T. marina* the transcriptome
904 assembly was first translated in all six open reading frames using the *ab initio* gene prediction
905 program GeneMarkS-T using the default parameters (Tang et al. 2015) and the longest resulting
906 predicted protein sequences were used for the searches. Forward hits meeting an e-value cutoff of
907 0.05 were subject to reciprocal BLASTp analyses against the *Homo sapiens* and *M. exilis* predicted
908 proteomes as well as the NCBI nr database. Any absent components were also subject to additional
909 tBLASTn searches in the nucleotide scaffolds. Hits were deemed positive if both forward hits and
910 at least two of three reciprocal BLAST results retrieved the correct orthologue with an e-value
911 cutoff of 0.05. Paralogous gene families were subject to further phylogenetic analyses.

912 Phylogenetic analyses were undertaken for the large, medium and small subunits from identified
913 AP1-4 and CopI families, as well as Syntaxin16 and Syntaxin5. Resolved backbone alignments
914 from the dataset curated for Karnkowska et al. (2019) were used and sequences from *T. marina*,
915 *P. pyriformis*, *B. nauphoetae*, and *S. strix* were iteratively aligned with the backbone alignment
916 using the profile option available through MUSCLE v3.8.31 (Edgar 2004). All alignments were
917 subsequently inspected and manually trimmed using Mesquite v. 3.5
918 (https://www.mesquiteproject.org/) to remove heterogeneous regions lacking obvious homology
919 and partial or misaligned sequences.

920 Maximum likelihood analyses were carried out using RAxML-HPC2 on XSEDE v.8.2.10 for non-
921 parametric bootstrap replicates (Stamatakis 2006). The best protein matrix model for RAxML was
922 tested using ProtTest v.3.4.2 (Abascal et al. 2005), set to account for Gamma rate variation (+G),
923 invariant sites (+I), and observed frequency of amino acids (+F) with default tree and a consensus
924 tree was obtained using the Consense package, available through the Phylip v.3.66 (Baum 1989).

925 Bayesian inference was carried out using MRBAYES on XSEDE v3.2.6 (Huelsenbeck and
926 Ronquist 2001). Parameters specified included 10 million Markov Chain Monte Carlo generations
927 under a mixed amino acid model with the number of gamma rate categories set a 4. The sampling
928 frequency was set to occur every 1000 generations with a burn-in of 0.25. Tree convergence was
929 ensured with the average standard deviation of split frequency values below 0.01. Both RAxML
930 and MRBAYES analyses were performed on the CIPRES webserver. RAxML bootstrap values as
931 well as Bayesian posterior probabilities were overlaid on the best Bayes topology with combined
932 values of >50 and >0.80, respectively, indicating branch support. All alignments are available upon
933 request.

934 **Prediction of proteins containing Fe-S clusters**

935 Fe-S cluster containing proteins were predicted with MetalPredator webserver (Valasatava et al.
936 2016). Predicted proteins were functionally annotated by BLAST searches against the NCBI nr
937 database and by InterProScan against the InterPro database (Blum et al. 2021). KEGG categories
938 were assigned by GhostKOALA searches against the KEGG database (Kanehisa et al. 2016).
939 Orthologous groups were created with the OrthoFinder 2 software (Emms and Kelly 2019) and the
940 Venn diagram of OG sharing among Preaxostyla was visualized using InteractiVenn.

941 **Search for mitochondrial proteins**

942 The comprehensive search for putative mitochondrial protein was performed for all five
943 Preaxostyla species, including previously analyzed *M. exilis* (Karnkowska et al. 2016) as a control.
944 Predicted proteins of each species were independently searched for similarity to known
945 mitochondrial/MRO-related proteins and the presence of mitochondrial signature sequences. The
946 general design of the search followed the previously described methodology (Karnkowska et al.
947 2016). Briefly, a custom mitochondrial protein sequence database was compiled using the
948 MitoMiner v4.0 database (Smith et al. 2012; Smith and Robinson 2018) as the core. Experimentally
949 confirmed mitochondrial proteins (at least one GFP tagging experiment or three different MS
950 experiments) coming from *H. sapiens*, *Mus musculus*, *Rattus norvegicus*, *Danio rerio*,
951 *Saccharomyces cerevisiae*, and *Schizosaccharomyces pombe* were used and supplemented with the
952 MROs proteins from sixteen different organisms (Stechmann et al. 2008; Mi-ichi et al. 2009;
953 Barberà et al. 2010; Alcock et al. 2012; Stairs et al. 2014; Noguchi et al. 2015; Nývltová et al.
954 2015; Leger et al. 2016; Leger et al. 2017; Pyrihová et al. 2018). Redundant homologues (90%
955 similarity threshold) were removed from the database using CD-HIT. The resulting non-redundant
956 database contained 6,979 proteins. Reciprocal BLAST analysis was performed for each predicted
957 proteome of Preaxostyla species with an e-value threshold of 0.001.

958 HMM searches using HMMER 3.1b2 and profile HMMs previously employed by Karnkowska et
959 al. (2016) were performed to specifically identify proteins involved in protein import and
960 translocation, as these were shown to be often divergent (Leger et al. 2016). Mitochondrial
961 targeting signals were detected using TargetP v1.1 (Emanuelsson et al. 2007) and MitoFates v1.1
962 (Fukasawa et al. 2015). Both programs indicate a probability of mitochondrial localization of the
963 protein. Hits with the probability of mitochondrial localization indicated to be >0.5 by both

964 programs were considered for manual verification. To find tail-anchored proteins, transmembrane
965 domains (TMDs) were predicted using TMHMM2.0 (Krogh et al. 2001) for all analyzed proteins.
966 Hits with a TMD within 32 amino acid residues from the C-terminus were kept for verification.
967 The mitochondrial β-barrel outer membrane proteins (MBOMPs) search has been conducted using
968 the pipeline described by Imai et al. (2011). The pipeline firstly identifies a β-signal
969 ($P_oxGh_yxH_yxH_y$ motif), required for the insertion into the membrane, in the C-terminus of the query
970 protein. Subsequently, the secondary structure of the stretch of 300 amino acid residues preceding
971 the β-signal is analyzed using PSIPRED (McGuffin et al. 2000) to check for a typical β-structure.
972 Significant hits, with at least 25% of the sequence predicted to form β-strands, no more than 10%
973 assumed by an α-helical structure, and no more than 50% of the eight residues of the β-signal
974 predicted as an α-helical structure, were analyzed further.

975 Any hits received by the methods described above were considered candidates and were further
976 inspected as follows. All candidates were BLAST-searched against NCBI-nr and the best hits with
977 the descriptions not including the terms 'low quality protein', 'hypothetical', 'unknown', etc. were
978 kept. For each hit, the Gene Ontology categories were assigned using InterProScan-5.36-75.0. If
979 the annotations received from BLAST or InterProScan corresponded to the originally suggested
980 function, the candidates were considered as verified. Otherwise they were considered as false
981 positives.

982 ***Trypanosoma brucei* mitoproteome-guided comparative analyses**

983 Predicted proteomes of *T. marina*, *P. pyriformis*, *S. strix*, *B. nauphoetae,* and *M. exilis* were
984 individually reverse BLAST-searched against the proteome of *T. brucei* (downloaded from
985 Tritrypdb.org; November 2019; Aslett et al. 2010). Only reciprocal best BLAST hits that were
986 identified in any of the previously published mitochondrial proteomes of *T. brucei* (Panigrahi et al.

987 2009; Peikert et al. 2017) were subjected to further phylogenetic analysis. For each such protein,
988 the data set for the tree construction was composed of hits from a custom BLAST database of
989 selected protist proteomes (downloaded from UniProt, November 2019; Bateman 2019). Protein
990 sequence sets were automatically aligned with MAFFT v7.453 using the L-INS-i refinement and a
991 maximum of 1,000 iterations, followed by trimming of sites with >70% gaps. ML trees
992 (Supplementary file 10) were inferred by IQ-TREE v 1.6.12 using the Posterior Mean Site
993 Frequency (PMSF) empirical model with the ultrafast bootstrapping strategy (1,000 replicates) and
994 a LG4X guide tree. Subcellular targeting of all proteins in the tree was predicted by using TargetP-
995 1.1; the presence of a signal peptide, a chloroplast targeting peptide or a mitochondrial targeting
996 peptide in the respective proteins is marked by the letters S, C, or M, respectively, at the very
997 beginning of the sequence names.

ACKNOWLEDGEMENTS

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No. 771592) and the Centre for research of pathogenicity and virulence of parasites (registration no. CZ.02.1.01/0.0/0.0/16_019/0000759). Computational resources were supplied by the project "e-Infrastruktura CZ" (e-INFRA LM2018140) provided within the program Projects of Large Research, Development and Innovations Infrastructures. Research in Karnkowska lab is supported by EMBO Installation Grant 4150 and Ministry of Education and Science, Poland and the Interdisciplinary Centre for Mathematical and Computational Modelling (ICM) University of Warsaw under computational allocation no G 72-16. Research in the Dacks Lab is supported by grants from the Natural Sciences and Research Council of Canada (RES0021028, RES0043758, and RES0046091) and SVP was supported by an Alberta Innovates Graduate Studentship (Doctoral) and Canadian Institutes of Health Research Canada Graduate Scholarships. ME was supported by the Czech Science Foundation project 22-29633S.

1012 **REFERENCES**

1013 Abascal F, Zardoya R, Posada D. 2005. ProtTest: selection of best-fit models of protein
1014 evolution. *Bioinformatics* 21:2104–2105.
1015 Agbu SO, Liang Y, Liu A, Anderson K V. 2018. The small GTPase RSG1 controls a final step in
1016 primary cilia initiation. *J Cell Biol* 217:413–427.
1017 Ahat E, Li J, Wang Y. 2019. New Insights Into the Golgi Stacking Proteins. *Front Cell Dev Biol*
1018 7:131.
1019 Alcock F, Webb CT, Dolezal P, Hewitt V, Shingu-Vasquez M, Likić VA, Traven A, Lithgow T.
1020 2012. A small Tim homohexamer in the relict mitochondrion of *Cryptosporidium*. *Mol Biol*
1021 *Evol* 29:113–122.
1022 Anderson IJ, Loftus BJ. 2005. *Entamoeba histolytica*: Observations on metabolism based on the
1023 genome sequence. *Exp Parasitol* 110:173–177.
1024 Andreini C, Banci L, Rosato A. 2016. Exploiting Bacterial Operons to Illuminate Human Iron-
1025 Sulfur Proteins. *J Proteome Res* 15:1308–1322.
1026 Andrew S. 2010. FastQC: a quality control tool for high throughput sequence data. Available
1027 online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc.
1028 Aridor M. 2022. A tango for coats and membranes: New insights into ER-to-Golgi traffic. *Cell*
1029 *Rep* 38:110258.
1030 Aslett M, Aurrecoechea C, Berriman M, Brestelli J, Brunk BP, Carrington M, Depledge DP,
1031 Fischer S, Gajria B, Gao X, et al. 2010. TriTrypDB: a functional genomic resource for the
1032 Trypanosomatidae. *Nucleic Acids Res* 38:D457–D462.
1033 Babady NE, Pang YP, Elpeleg O, Isaya G. 2007. Cryptic proteolytic activity of dihydrolipoamide
1034 dehydrogenase. *Proc Natl Acad Sci U S A* 104:6158–6163.
1035 Barberà MJ, Ruiz-Trillo I, Tufts JY, Bery A, Silberman JD, Roger AJ. 2010. *Sawyeria*
1036 *marylandensis* (Heterolobosea) has a hydrogenosome with novel metabolic properties.
1037 *Eukaryot Cell* 9:1913–1924.
1038 Barlow LD, Nývltová E, Aguilar M, Tachezy J, Dacks JB. 2018. A sophisticated, differentiated
1039 Golgi in the ancestor of eukaryotes. *BMC Biol* 16:27.
1040 Bateman A. 2019. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res* 47:D506–
1041 D515.
1042 van der Bliek AM, Shen Q, Kawajiri S. 2013. Mechanisms of mitochondrial fission and fusion.
1043 *Cold Spring Harb Perspect Biol* 5(6): a011072.
1044 Blum M, Chang HY, Chuguransky S, Grego T, Kandasaamy S, Mitchell A, Nuka G, Paysan-
1045 Lafosse T, Qureshi M, Raj S, et al. 2021. The InterPro protein families and domains
1046 database: 20 years on. *Nucleic Acids Res* 49:D344–D354.
1047 Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence
1048 data. *Bioinformatics* 30:2114–2120.
1049 Boncompain G, Weigel A v. 2018. Transport and sorting in the Golgi complex: multiple
1050 mechanisms sort diverse cargo. *Curr Opin Cell Biol* 50:94–101.
1051 Borgese N, Brambillasca S, Colombo S. 2007. How tails guide tail-anchored proteins to their
1052 destinations. *Curr Opin Cell Biol* 19:368–375.
1053 Braymer JJ, Freibert SA, Rakwalska-Bange M, Lill R. 2021. Mechanistic concepts of iron-sulfur
1054 protein biogenesis in Biology. *Biochim Biophys Acta Mol Cell Res* 1868(1):118863.
1055 Buchfink B, Xie C, Huson DH. 2014. Fast and sensitive protein alignment using DIAMOND.
1056 *Nature Methods* 12:59–60.

1057 Butenko A, Opperdoes FR, Flegontova O, Horák A, Hampl V, Keeling P, Gawryluk RMR,
1058 Tikhonenkov D, Flegontov P, Lukeš J. 2020. Evolution of metabolic capabilities and
1059 molecular features of diplonemids, kinetoplastids, and euglenids. *BMC Biol* 18:1–28.
1060 Capella-Gutiérrez S, Silla-Martínez JM, Gabaldón T. 2009. trimAl: A tool for automated
1061 alignment trimming in large-scale phylogenetic analyses. *Bioinformatics* 25:1972–1973.
1062 Carlton JM, Hirt RP, Silva JC, Delcher AL, Schatz M, Zhao Q, Wortman JR, Bidwell SL,
1063 Alsmark UCM, Besteiro S, et al. 2007. Draft genome sequence of the sexually transmitted
1064 pathogen *Trichomonas vaginalis*. *Science* 315:207–212.
1065 Christensen GA, Wymore AM, King AJ, Podar M, Hurt RA, Santillan EU, Soren A, Brandt CC,
1066 Brown SD, Palumbo A v., et al. 2016. Development and validation of broad-range
1067 qualitative and clade-specific quantitative molecular probes for assessing mercury
1068 methylation in the environment. *Appl Environ Microbiol* 82:6068–6078.
1069 Cooper CJ, Zheng K, Rush KW, Johs A, Sanders BC, Pavlopoulos GA, Kyrpides NC, Podar M,
1070 Ovchinnikov S, Ragsdale SW, et al. 2020. Structure determination of the HgcAB complex
1071 using metagenome sequence data: insights into microbial mercury methylation.
1072 *Communications Biology* 3:1–9.
1073 Criscuolo A, Gribaldo S. 2010. BMGE (Block Mapping and Gathering with Entropy): a new
1074 software for selection of phylogenetic informative regions from multiple sequence
1075 alignments. *BMC Evol Biol* 10:210.
1076 Dean S, Sunter JD, Wheeler RJ. 2017. TrypTag.org: A Trypanosome Genome-wide Protein
1077 Localisation Resource. *Trends Parasitol* 33:80–82.
1078 Denic V. 2012. A portrait of the GET pathway as a surprisingly complicated young man. *Trends*
1079 *Biochem Sci* 37:411–417.
1080 Derelle R, Torruella G, Klimeš V, Brinkmann H, Kim E, Vlček Č, Lang BF, Eliáš M. 2015.
1081 Bacterial proteins pinpoint a single eukaryotic root. *Proc Natl Acad Sci U S A* 112:E693–
1082 E699.
1083 Dewees SI, Vargová R, Hardin KR, Turn RE, Devi S, Linnert J, Wolfrum U, Caspary T, Eliáš M,
1084 Kahn RA. 2022. Phylogenetic profiling and cellular analyses of ARL16 reveal roles in
1085 traffic of IFT140 and INPP5E. *Mol Biol Cell* 33:ar33.
1086 Diamond LS. 1982. A new liquid medium for xenic cultivation of *Entamoeba histolytica* and
1087 other lumen-dwelling protozoa. *J Parasitol* 68:958–959.
1088 Dolezal P, Likic V, Tachezy J, Lithgow T. 2006. Evolution of the Molecular Machines for
1089 Protein Import into Mitochondria. *Science* 313:314–319.
1090 Ducker GS, Rabinowitz JD. 2017. Cell Metabolism Review One-Carbon Metabolism in Health
1091 and Disease. *Cell Metab* 25:27–42.
1092 Duschak VG, Cazzulo JJ. 1991. Subcellular localization of glutamate dehydrogenases and
1093 alanine aminotransferase in epimastigotes of *Trypanosoma cruzi*. *FEMS Microbiol Lett*
1094 83:131–135.
1095 Eberhardt EL, Ludlam A v., Tan Z, Cianfrocco MA. 2020. Miro: A molecular switch at the
1096 center of mitochondrial regulation. *Protein Sci* 29:1269–1284.
1097 Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7(10): e1002195.
1098 Edgar RC. 2004. MUSCLE: multiple sequence alignment with high accuracy and high
1099 throughput. *Nucleic Acids Res* 32:1792–1797.
1100 Emanuelsson O, Brunak S, von Heijne G, Nielsen H. 2007. Locating proteins in the cell using
1101 TargetP, SignalP and related tools. *Nat Protoc* 2:953–971.

1102 Emms DM, Kelly S. 2019. OrthoFinder: Phylogenetic orthology inference for comparative
1103 genomics. *Genome Biol* 20:238.
1104 Felsenstein J. 1989. PHYLIP: Phylogeny Inference Package (Version 3.2). *Cladistics* 5:164-166.
1105 Fukasawa Y, Tsuji J, Fu SC, Tomii K, Horton P, Imai K. 2015. MitoFates: Improved prediction
1106 of mitochondrial targeting sequences and their cleavage sites. *Molecular and Cellular*
1107 *Proteomics* 14:1113–1126.
1108 Gentekaki E, Curtis BA, Stairs CW, Klimeš V, Eliáš M, Salas-Leiva DE, Herman EK, Eme L,
1109 Arias MC, Henrissat B, et al. 2017. Extreme genome diversity in the hyper-prevalent
1110 parasitic eukaryote *Blastocystis*. *PLoS Biol* 15(9):e2003769.
1111 Gilmour CC, Elias DA, Kucken AM, Brown SD, Palumbo A v., Schadt CW, Wall JD. 2011.
1112 Sulfate-reducing bacterium *Desulfovibrio desulfuricans* ND132 as a model for
1113 understanding bacterial mercury methylation. *Appl Environ Microbiol* 77:3938–3951.
1114 Greening C, Biswas A, Carere CR, Jackson CJ, Taylor MC, Stott MB, Cook GM, Morales SE.
1115 2016. Genomic and metagenomic surveys of hydrogenase distribution indicate H 2 is a
1116 widely utilised energy source for microbial growth and survival. *ISME Journal* 10:761–777.
1117 Haas BJ, Delcher AL, Mount SM, Wortman JR, Smith RK, Hannick LI, Maiti R, Ronning CM,
1118 Rusch DB, Town CD, et al. 2003. Improving the *Arabidopsis* genome annotation using
1119 maximal transcript alignment assemblies. *Nucleic Acids Res* 31:5654–5666.
1120 Haas BJ, Salzberg SL, Zhu W, Pertea M, Allen JE, Orvis J, White O, Robin CR, Wortman JR.
1121 2008. Automated eukaryotic gene structure annotation using EVidenceModeler and the
1122 Program to Assemble Spliced Alignments. *Genome Biol* 9:R7.
1123 Haddad I, Hiller K, Frimmersdorf E, Benkert B, Schomburg D, Jahn D. 2009. An emergent self-
1124 organizing map based analysis pipeline for comparative metabolome studies. *In Silico Biol*
1125 9:163–178.
1126 Hampl V. 2017. Preaxostyla. In: Handbook of the Protists. Cham: Springer International
1127 Publishing. p. 1139–1174.
1128 Hampl V, Silberman JD, Stechmann A, Diaz-Triviño S, Johnson PJ, Roger AJ. 2008. Genetic
1129 Evidence for a Mitochondriate Ancestry in the 'Amitochondriate' Flagellate *Trimastix*
1130 *pyriformis*. *PLoS One* 3(1):e1383.
1131 Hazelwood LA, Daran JM, van Maris AJA, Pronk JT, Dickinson JR. 2008. The Ehrlich pathway
1132 for fusel alcohol production: A century of research on *Saccharomyces cerevisiae*
1133 metabolism. *Appl Environ Microbiol* 74:2259–2266.
1134 Heberle H, Meirelles VG, da Silva FR, Telles GP, Minghim R. 2015. InteractiVenn: a web-based
1135 tool for the analysis of sets through Venn diagrams. *BMC Bioinformatics* 16:169.
1136 Hirst J, Schlacht A, Norcott JP, Traynor D, Bloomfield G, Antrobus R, Kay RR, Dacks JB,
1137 Robinson MS. 2014. Characterization of TSET, an ancient and widespread membrane
1138 trafficking complex. *Elife* 3:e02866.
1139 Huelsenbeck JP, Ronquist F. 2001. MRBAYES: Bayesian inference of phylogenetic trees.
1140 *Bioinformatics* 17:754–755.
1141 Igamberdiev AU, Bykova N v., Ens W, Hill RD. 2004. Dihydrolipoamide dehydrogenase from
1142 porcine heart catalyzes NADH-dependent scavenging of nitric oxide. *FEBS Lett* 568:146–
1143 150.
1144 Imai K, Fujita N, Gromiha MM, Horton P. 2011. Eukaryote-wide sequence analysis of
1145 mitochondrial β-barrel outer membrane proteins. *BMC Genomics* 12:79.

1146 Jerlström-Hultqvist J, Einarsson E, Xu F, Hjort K, Ek B, Steinhauf D, Hultenby K, Bergquist J,
1147 Andersson JO, Svärd SG. 2013. Hydrogenosomes in the diplomonad *Spironucleus*
1148 *salmonicida*. *Nat Commun* 4:2493.
1149 Kanehisa M, Sato Y, Morishima K. 2016. BlastKOALA and GhostKOALA: KEGG Tools for
1150 Functional Characterization of Genome and Metagenome Sequences. *J Mol Biol*
1151 428(4):726-731.
1152 Karnkowska A, Treitli SC, Brzoň O, Novák L, Vacek V, Soukal P, Barlow LD, Herman EK,
1153 Pipaliya S V, Pánek T, et al. 2019. The oxymonad genome displays canonical eukaryotic
1154 complexity in the absence of a mitochondrion. *Mol Biol Evol* 36:2292–2312.
1155 Karnkowska A, Vacek V, Zubáčová Z, Treitli SC, Petrželková R, Eme L, Novák L, Žárský V,
1156 Barlow LD, Herman EK, et al. 2016. A eukaryote without a mitochondrial organelle.
1157 *Current Biology* 26:1274–1284.
1158 Katoh K, Standley DM. 2013. MAFFT multiple sequence alignment software version 7:
1159 improvements in performance and usability. *Mol Biol Evol* 30:772–780.
1160 Klinger CM, Karnkowska A, Herman EK, Hampl V, Dacks JB. 2016. Phylogeny and Evolution.
1161 In: Molecular Parasitology. Vienna: Springer Vienna. p. 383–408.
1162 Klinger CM, Klute MJ, Dacks JB. 2013. Comparative genomic analysis of multi-subunit
1163 tethering complexes demonstrates an ancient pan-eukaryotic complement and sculpting in
1164 Apicomplexa. *PLoS One* 8:e76278.
1165 Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and
1166 accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome*
1167 *Res* 27:722–736.
1168 Krogh A, Larsson B, Von Heijne G, Sonnhammer ELL. 2001. Predicting transmembrane protein
1169 topology with a hidden Markov model: Application to complete genomes. *J Mol Biol*
1170 305:567–580.
1171 Kulkarni-Gosavi P, Makhoul C, Gleeson PA. 2019. Form and function of the Golgi apparatus:
1172 scaffolds, cytoskeleton and signalling. *FEBS Lett* 593:2289–2305.
1173 Langousis G, Cavadini S, Boegholm N, Lorentzen E, Kempf G, Matthias P. 2022. Structure of
1174 the ciliogenesis-associated CPLANE complex. *Sci Adv* 8:eabn0832.
1175 Lee I, Tiwari N, Dunlop MH, Graham M, Liu X, Rothman JE. 2014. Membrane adhesion dictates
1176 Golgi stacking and cisternal morphology. *Proc Natl Acad Sci U S A* 111:1849–1854.
1177 Leger MM, Eme L, Hug LA, Roger AJ. 2016. Novel hydrogenosomes in the microaerophilic
1178 jakobid *Stygiella incarcerata*. *Mol Biol Evol* 33:2318–2336.
1179 Leger MM, Kolisko M, Kamikawa R, Stairs CW, Kume K, Čepička I, Silberman JD, Andersson
1180 JO, Xu F, Yabuki A, et al. 2017. Organelles that illuminate the origins of *Trichomonas*
1181 hydrogenosomes and Giardia mitosomes. *Nat Ecol Evol* 1:0092.
1182 Li J, Ahat E, Wang Y. 2019. Golgi Structure and Function in Health, Stress, and Diseases.
1183 *Results Probl Cell Differ* 67:441–485.
1184 Li W, Godzik A. 2006. Cd-hit: A fast program for clustering and comparing large sets of protein
1185 or nucleotide sequences. *Bioinformatics* 22:1658–1659.
1186 Liapounova NA, Hampl V, Gordon PMK, Sensen CW, Gedamu L, Dacks JB. 2006.
1187 Reconstructing the mosaic glycolytic pathway of the anaerobic eukaryote
1188 *Monocercomonoides*. *Eukaryot Cell* 5:2138–2146.
1189 Lindmark DG, Müller M. 1973. Hydrogenosome, a cytoplasmic organelle of the anaerobic
1190 flagellate *Tritrichomonas foetus*, and its role in pyruvate metabolism. *J Biol Chem*
1191 248:7724–7728.

1192 Loman NJ, Quick J, Simpson JT. 2015. A complete bacterial genome assembled de novo using
1193 only nanopore sequencing data. *Nat Methods* 12:733–735.
1194 Lucattini R, Likić VA, Lithgow T. 2004. Bacterial Proteins Predisposed for Targeting to
1195 Mitochondria. *Mol Biol Evol* 21:652–658.
1196 McGuffin LJ, Bryson K, Jones DT. 2000. The PSIPRED protein structure prediction server.
1197 *Bioinformatics* 16:404–405.
1198 Mi-ichi F, Yousuf MA, Nakada-Tsukui K, Nozaki T, Abu Yousuf M, Nakada-Tsukui K, Nozaki
1199 T. 2009. Mitosomes in *Entamoeba histolytica* contain a sulfate activation pathway.
1200 *Proceedings of the National Academy of Sciences* 106:21731–21736.
1201 Miller MA, Pfeiffer W, Schwartz T. 2010. Creating the CIPRES Science Gateway for inference
1202 of large phylogenetic trees. In: 2010 Gateway Computing Environments Workshop (GCE).
1203 IEEE. p. 1–8.
1204 Moriya Y, Itoh M, Okuda S, Yoshizawa AC, Kanehisa M. 2007. KAAS: an automatic genome
1205 annotation and pathway reconstruction server. *Nucleic Acids Res* 35:W182–W185.
1206 Müller M, Mentel M, van Hellemond JJ, Henze K, Woehle C, Gould SB, Yu R-Y, van der
1207 Giezen M, Tielens AGM, Martin WF. 2012. Biochemistry and evolution of anaerobic
1208 energy metabolism in eukaryotes. *Microbiol Mol Biol Rev* 76:444–495.
1209 Nakjang S, Williams TA, Heinz E, Watson AK, Foster PG, Sendra KM, Heaps SE, Hirt RP,
1210 Embley TM. 2013. Reduction and expansion in microsporidian genome evolution: New
1211 insights from comparative genomics. *Genome Biol Evol* 5:2285–2303.
1212 Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: A Fast and Effective
1213 Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies. *Mol Biol Evol*
1214 32:268–274.
1215 Noguchi F, Shimamura S, Nakayama T, Yazaki E, Yabuki A, Hashimoto T, Inagaki Y, Fujikura
1216 K, Takishita K. 2015. Metabolic Capacity of Mitochondrion-related Organelles in the Free-
1217 living Anaerobic Stramenopile *Cantina marsupialis*. *Protist* 166:534–550.
1218 Novák L, Zubáčová Z, Karnkowska A, Kolisko M, Hroudová M, Stairs CW, Simpson AGB,
1219 Keeling PJ, Roger AJ, Čepička I, et al. 2016. Arginine deiminase pathway enzymes:
1220 evolutionary history in metamonads and other eukaryotes. *BMC Evol Biol* 16:197.
1221 Nývltová E, Stairs CW, Hrdý I, Rídl J, Mach J, Paɥes J, Roger AJ, Tachezy J. 2015. Lateral gene
1222 transfer and gene duplication played a key role in the evolution of *Mastigamoeba balamuthi*
1223 hydrogenosomes. *Mol Biol Evol* 32:1039–1055.
1224 Nývltová E, Šut'ák R, Žárský V, Harant K, Hrdý I, Tachezy J. 2017. Lateral gene transfer of p-
1225 cresol- and indole-producing enzymes from environmental bacteria to *Mastigamoeba*
1226 *balamuthi. Environ Microbiol* 19:1091–1102.
1227 O'Kelly CJ, Farmer MA, Nerad TA. 1999. Ultrastructure of *Trimastix pyriformis* (Klebs)
1228 Bernard et al.: similarities of Trimastix species with retortamonad and jakobid flagellates.
1229 *Protist* 150:149–162.
1230 Onuț-Brännström I, Stairs CW, Campos KIA, Thorén MH, Ettema TJG, Keeling PJ, Bass D,
1231 Burki F. 2023. A Mitosome With Distinct Metabolism in the Uncultured Protist Parasite
1232 *Paramikrocytos canceri* (Rhizaria, Ascetosporea). *Genome Biol Evol* 15(3):evad022.
1233 Panigrahi AK, Ogata Y, Zíková A, Anupama A, Dalley RA, Acestor N, Myler PJ, Stuart KD.
1234 2009. A comprehensive analysis of *Trypanosoma brucei* mitochondrial proteome.
1235 *Proteomics* 9:434–450.
1236 Park K, Ju S, Kim N, Park SY. 2021. The Golgi complex: a hub of the secretory pathway. *BMB*
1237 *Rep* 54:246.

1238 Parks DH, Chuvochina M, Rinke C, Mussig AJ, Chaumeil PA, Hugenholtz P. 2022. GTDB: an
1239 ongoing census of bacterial and archaeal diversity through a phylogenetically consistent,
1240 rank normalized and complete genome-based taxonomy. *Nucleic Acids Res* 50:D785–D794.
1241 Payne MJ, Chapman A, Cammack R. 1993. Evidence for an [Fe]-type hydrogenase in the
1242 parasitic protozoan *Trichomonas vaginalis*. *FEBS Lett* 317:101–104.
1243 Peikert CD, Mani J, Morgenstern M, Käser S, Knapp B, Wenger C, Harsman A, Oeljeklaus S,
1244 Schneider A, Warscheid B. 2017. Charting organellar importomes by quantitative mass
1245 spectrometry. *Nat Commun* 8:15272.
1246 Peña-Diaz J, Montalvetti A, Flores C-L, Constán A, Hurtado-Guerrero R, De Souza W, Gancedo
1247 C, Ruiz-Perez LM, Gonzalez-Pacanowska D. 2004. Mitochondrial localization of the
1248 mevalonate pathway enzyme 3-Hydroxy-3-methyl-glutaryl-CoA reductase in the
1249 Trypanosomatidae. *Mol Biol Cell* 15:1356–1363.
1250 Peña-Diaz P, Braymer JJ, Vacek V, Zelená M, Lometto S, Hrdý I, Treitli SC, Hochberg G, Py B,
1251 Lill R, et al. 2023. Characterisation of the SUF FeS cluster machinery in the amitochondriate
1252 eukaryote *Monocercomonoides exilis*. *bioRxiv* 2023.03.30.534840.
1253 Petrat F, Paluch S, Dogruöz E, Dörfler P, Kirsch M, Korth HG, Sustmann R, de Groot H. 2003.
1254 Reduction of Fe(III) Ions Complexed to Physiological Ligands by Lipoyl Dehydrogenase
1255 and Other Flavoenzymes in Vitro: Implications for an Enzymatic Reduction of Fe(III) Ions
1256 of the Labile Iron Pool. *Journal of Biological Chemistry* 278:46403–46413.
1257 Pfanner N, Warscheid B, Wiedemann N. 2019. Mitochondrial proteins: from biogenesis to
1258 functional networks. *Nature Reviews Molecular Cell Biology* 20:267–284.
1259 Podar M, Gilmour CC, Brandt CC, Soren A, Brown SD, Crable BR, Palumbo A v., Somenahally
1260 AC, Elias DA. 2015. Global prevalence and distribution of genes and microorganisms
1261 involved in mercury methylation. *Sci Adv* 1(9):e1500675.
1262 Poinar GO. 2009. Description of an early Cretaceous termite (Isoptera: Kalotermitidae) and its
1263 associated intestinal protozoa, with comments on their co-evolution. *Parasit Vectors* 2:12.
1264 Price MN, Dehal PS, Arkin AP. 2010. FastTree 2 – Approximately Maximum-Likelihood Trees
1265 for Large Alignments. *PLoS One* 5:e9490.
1266 Pyrihová E, Motyčková A, Voleman L, Wandyszewska N, Fišer R, Seydlová G, Roger A,
1267 Kolísko M, Doležal P. 2018. A single tim translocase in themitosomes of *Giardia*
1268 *intestinalis* illustrates convergence of protein import machines in anaerobic eukaryotes.
1269 *Genome Biol Evol* 10:2813–2822.
1270 Rada P, Makki A, Žárský V, Tachezy J. 2019. Targeting of tail-anchored proteins to
1271 *Trichomonas vaginalis* hydrogenosomes. *Mol Microbiol* 111:588–603.
1272 Richter DJ, Berney C, Strassert JFH, Poh Y-P, Herman EK, Muñoz-Gómez SA, Wideman JG,
1273 Burki F, Vargas C de. 2022. EukProt: a database of genome-scale predicted proteins across
1274 the diversity of eukaryotes. *bioRxiv* 2020.06.30.180687.
1275 Saas J, Ziegelbauer K, von Haeseler A, Fast B, Boshart M. 2000. A developmentally regulated
1276 aconitase related to iron-regulatory protein-1 is localized in the cytoplasm and in the
1277 mitochondrion of *Trypanosoma brucei*. *J Biol Chem* 275:2745–2755.
1278 Sakuraba H, Satomura T, Kawakami R, Yamamoto S, Kawarabayasi Y, Kikuchi H, Ohshima T.
1279 2002. L-aspartate oxidase is present in the anaerobic hyperthermophilic archaeon
1280 *Pyrococcus horikoshii* OT-3: Characteristics and role in the de novo biosynthesis of
1281 nicotinamide adenine dinucleotide proposed by genome sequencing. *Extremophiles* 6:275–
1282 281.

1283 Salas-Leiva DE, Tromer EC, Curtis BA, Jerlström-Hultqvist J, Kolisko M, Yi Z, Salas-Leiva JS,
1284 Gallot-Lavallée L, Williams SK, Kops GJPL, et al. 2021. Genomic analysis finds no
1285 evidence of canonical eukaryotic DNA processing complexes in a free-living protist. *Nat*
1286 *Commun* 12:6003.
1287 Schut GJ, Adams MWW. 2009. The iron-hydrogenase of *Thermotoga maritima* utilizes
1288 ferredoxin and NADH synergistically: A new perspective on anaerobic hydrogen
1289 production. *J Bacteriol* 191:4451–4457.
1290 Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E v., Zdobnov EM. 2015. BUSCO:
1291 assessing genome assembly and annotation completeness with single-copy orthologs.
1292 31:3210–3212.
1293 Smith AC, Blackshaw JA, Robinson AJ. 2012. MitoMiner: A data warehouse for mitochondrial
1294 proteomics data. *Nucleic Acids Res* 40:D1160-7.
1295 Smith AC, Robinson AJ. 2018. MitoMiner v4.0: an updated database of mitochondrial
1296 localization evidence, phenotypes and diseases. *Nucleic Acids Res* 47 D1225–D1228.
1297 Song L, Shankar DS, Florea L. 2016. Rascaf: Improving Genome Assembly with RNA
1298 Sequencing Data. *Plant Genome* 9:0.
1299 Spalding MD, Prigge ST. 2010. Lipoic acid metabolism in microbial pathogens. *Microbiol Mol*
1300 *Biol Rev* 74:200–228.
1301 Stairs CW, Eme L, Brown MW, Mutsaers C, Susko E, Dellaire G, Soanes DM, van der Giezen
1302 M, Roger AJ. 2014. A SUF Fe-S cluster biogenesis system in the mitochondrion-related
1303 organelles of the anaerobic protist *Pygsuia*. *Current Biology* 24:1176–1186.
1304 Stairs CW, Leger MM, Roger AJ. 2015. Diversity and origins of anaerobic metabolism in
1305 mitochondria and related organelles. *Philos Trans R Soc Lond B Biol Sci* 370:20140326.
1306 Stairs CW, Táborský P, Salomaki ED, Kolisko M, Pánek T, Eme L, Hradilová M, Vlček Č,
1307 Jerlström-Hultqvist J, Roger AJ, et al. 2021. Anaeramoebae are a divergent lineage of
1308 eukaryotes that shed light on the transition from anaerobic mitochondria to
1309 hydrogenosomes. *Curr Biol* 31:5605-5612.e5.
1310 Stamatakis A. 2006. The RAxML 7.0.4 Manual. *Bioinformatics* 22(21):2688–2690.
1311 Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of
1312 large phylogenies. *Bioinformatics* 30:1312–1313.
1313 Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron
1314 submodel. *Bioinformatics* 19:ii215–ii225.
1315 Stechmann A, Hamblin K, Pérez-Brocal V, Gaston D, Richmond GSS, Van Der Giezen M, Clark
1316 CGG, Roger AJ, Giezen M Van Der, Clark CGG, et al. 2008. Report Organelles in
1317 *Blastocystis* that Blur the Distinction between Mitochondria and Hydrogenosomes. *Current*
1318 *Biology* 18:580–585.
1319 Stephens TG, Bhattacharya D, Ragan MA, Chan CX. 2016. PhySortR: a fast, flexible tool for
1320 sorting phylogenetic trees in R. *PeerJ* 4:e2038.
1321 Stipanuk MH. 1986. Metabolism of sulfur-containing amino acids. *Annu Rev Nutr* 6:179-209.
1322 Swearingen JW, Fuentes DE, Araya MA, Plishker MF, Saavedra CP, Chasteen TG, Vásquez CC.
1323 2006. Expression of the ubiE gene of *Geobacillus stearothermophilus* V in *Escherichia coli*
1324 K-12 mediates the evolution of selenium compounds into the headspace of selenite- and
1325 selenate-amended cultures. *Appl Environ Microbiol* 72:963–967.
1326 Tachezy J, Doležal P. 2007. Iron–Sulfur Proteins and Iron–Sulfur Cluster Assembly in
1327 Organisms with Hydrogenosomes and Mitosomes. In: Origin of Mitochondria and
1328 Hydrogenosomes. Berlin, Heidelberg: Springer Berlin Heidelberg. p. 105–133.

1329 Tang S, Lomsadze A, Borodovsky M. 2015. Identification of protein coding regions in RNA
1330 transcripts. *Nucleic Acids Res* 43:1–10.
1331 Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to Identify Repetitive Elements in
1332 Genomic Sequences. In: Current Protocols in Bioinformatics. Vol. 25. Hoboken, NJ, USA:
1333 John Wiley & Sons, Inc. p. 4.10.1-4.10.14.
1334 Tice AK, Žihala D, Pánek T, Jones RE, Salomaki ED, Nenarokov S, Burki F, Eliáš M, Eme L,
1335 Roger AJ, et al. 2021. PhyloFisher: A phylogenomic package for resolving eukaryotic
1336 relationships. *PLoS Biol* 19(8):e3001365.
1337 Treitli SC, Kolisko M, Husník F, Keeling PJ, Hampl V. 2019. Revealing the metabolic capacity
1338 of *Streblomastix strix* and its bacterial symbionts using singlecell metagenomics. *Proc Natl*
1339 *Acad Sci U S A* 116:19675–19684.
1340 Treitli SC, Kotyk M, Yubuki N, Jirounková E, Vlasáková J, Smejkalová P, Šípek P, Čepička I,
1341 Hampl V. 2018. Molecular and Morphological Diversity of the Oxymonad Genera
1342 *Monocercomonoides* and *Blattamonas* gen. nov. *Protist* 169:744–783.
1343 Treitli SC, Peña-Diaz P, Hałakuc P, Karnkowska A, Hampl V. 2021. High quality genome
1344 assembly of the amitochondriate eukaryote *Monocercomonoides exilis*. *Microb Genom*
1345 7(12):000745.
1346 Tsaousis AD, Kunji ERS, Goldberg A v, Lucocq JM, Hirt RP, Embley TM. 2008. A novel route
1347 for ATP acquisition by the remnant mitochondria of *Encephalitozoon cuniculi*. *Nature*
1348 453:553–556.
1349 Vacek V, Novák LVF, Treitli SC, Táborský P, Čepička I, Kolísko M, Keeling PJ, Hampl V,
1350 Ruiz-Trillo I. 2018. Fe-S Cluster Assembly in Oxymonads and Related Protists. *Mol Biol*
1351 *Evol* 35(11):2712-2718.
1352 Valasatava Y, Rosato A, Banci L, Andreini C. 2016. MetalPredator: A web server to predict iron-
1353 sulfur cluster binding proteomes. *Bioinformatics* 32:2850–2852.
1354 Vargová R, Hanousková P, Salamonová J, Žihala D, Silberman JD, Eliáš M, Čepička I. 2022.
1355 Evidence for an Independent Hydrogenosome-to-Mitosome Transition in the CL3 Lineage
1356 of Fornicates. *Front Microbiol* 13:1229.
1357 Vargová R, Wideman JG, Derelle R, Klimeš V, Kahn RA, Dacks JB, Eliáš M. 2021. A
1358 Eukaryote-Wide Perspective on the Diversity and Evolution of the ARF GTPase Protein
1359 Family. *Genome Biol Evol* 13:evab157.
1360 Vlahou G, Eliáš M, von Kleist-Retzow JC, Wiesner RJ, Rivero F. 2011. The Ras related GTPase
1361 Miro is not required for mitochondrial transport in *Dictyostelium discoideum*. *Eur J Cell*
1362 *Biol* 90:342–355.
1363 Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q,
1364 Wortman J, Young SK, et al. 2014. Pilon: An integrated tool for comprehensive microbial
1365 variant detection and genome assembly improvement. *PLoS One* 9:e112963.
1366 Williams K, Lowe PN, Leadlay PF. 1987. Purification and characterization of pyruvate:
1367 ferredoxin oxidoreductase from the anaerobic protozoon *Trichomonas vaginalis*. *Biochem J*
1368 246:529–536.
1369 Williams SK, Hultqvist JJ, Eglit Y, Salas-Leiva DE, Curtis B, Orr R, Stairs CW, Simpson AGB,
1370 Roger AJ. 2023. Extreme mitochondrial reduction in a novel group of free-living
1371 metamonads. *bioRxiv* 2023.05.03.539051.
1372 Xia L, Björnstedt M, Nordman T, Eriksson LC, Olsson JM. 2001. Reduction of ubiquinone by
1373 lipoamide dehydrogenase. An antioxidant regenerating pathway. *Eur J Biochem* 268:1486–
1374 1490.

1375 Xu F, Jerlström-Hultqvist J, Einarsson E, Ástvaldsson Á, Svärd SG, Andersson JO. 2014. The
1376 Genome of *Spironucleus salmonicida* Highlights a Fish Pathogen Adapted to Fluctuating
1377 Environments. *PLoS Genet* 10:e1004053.
1378 Yagi T, Shounaka M, Yamamoto S. 1990. Distribution of aspartate aminotrasnsferase activity in
1379 yeasts, and purification and characterization of mitochondrial and cytosolic isoenzymes
1380 from *Rhodotorula marina*. *J Biochem* 107:151–159.
1381 Yazaki E, Kume K, Shiratori T, Eglit Y, Tanifuji G, Harada R, Simpson AGB, Ishida KI,
1382 Hashimoto T, Inagaki Y. 2020. Barthelonids represent a deep-branching metamonad clade
1383 with mitochondrion-related organelles predicted to generate no ATP. *Proceedings of the*
1384 *Royal Society B* 287:20201538.
1385 Zhang Q, Táborský P, Silberman JD, Pánek T, Čepička I, Simpson AGB. 2015. Marine Isolates
1386 of *Trimastix marina* Form a Plesiomorphic Deep-branching Lineage within Preaxostyla,
1387 Separate from Other Known Trimastigids (*Paratrimastix* n. gen.). *Protist* 166:468–491.
1388 Zhou X, Yuan Y, Yang Y, Rutzke M, Thannhauser TW, Kochian L v., Li L. 2009. Involvement
1389 of a broccoli COQ5 methyltransferase in the production of volatile selenium compounds.
1390 *Plant Physiol* 151:528–540.
1391 Zítek J, Füssy Z, Treitli SC, Peña-Diaz P, Vaitová Z, Zavadska D, Harant K, Hampl V. 2022.
1392 Reduced mitochondria provide an essential function for the cytosolic methionine cycle.
1393 *bioRxiv* 2022.04.01.486701.
1394 Zubáčová Z, Novák L, Bublíková J, Vacek V, Fousek J, Rídl J, Tachezy J, Doležal P, Vlček Č,
1395 Hampl V. 2013. The Mitochondrion-Like Organelle of *Trimastix pyriformis* Contains the
1396 Complete Glycine Cleavage System. *PLoS One* 8:e55417.
1397

1398 **FIGURE LEGENDS**

1399 **Figure 1.** Distribution of orthologous groups + singletons (OGs) and other discussed features
1400 among metamonads. **A)** Schematic phylogenetic tree of Preaxostyla within Metamonada (based on
1401 the topology of the phylogenomic tree reported by Stairs et al. (2021) showing the distribution of
1402 categories of OGs, lateral gene transfer events, and key biological features mapped to nodes. Core
1403 set: estimated relaxed core gene set for Metamonada; increment: clade-specific increment of the
1404 relaxed core gene set; specific: unique OGs for the external leaves (species); new: novel or
1405 extremely divergent genes detected in the species or lineage; LGT: lateral gene transfer events;
1406 SUF: iron-sulfur cluster assembly SUF system; L2: glycine cleavage system protein L2 (GCS-L2);
1407 FF: flavodoxin-ferredoxin fusion protein; FH: fused hydrogenase; FC: folate cycle; MC:
1408 methionine cycle; PotE: b(0,+)-type amino acid transporter; SPNS: sphingolipid transporter; RFC:
1409 Reduced Folate Carrier; TauE: sulfite exporter TauE/SafE family protein; SP: sugar porters; Hg:
1410 mercury methylation; Se: selenium volatilization; pC: *p*-cresol synthesis; ADI: arginine deiminase;
1411 ADIp: arginine deiminase pathway; CK: carbamate kinase. **B)** Venn diagram showing the
1412 distribution of OGs among the three main analyzed lineages of Metamonada. **C)** Venn diagram
1413 showing the distribution of Preaxostyla-specific OGs among the three lineages of Preaxostyla. **D)**
1414 Venn diagram showing the distribution of Oxymonadida-specific OGs among the three studied
1415 species of Oxymonadida. The IDs and functional annotation of OGs is given in Supplementary file
1416 2. The IDs and functional annotation of LGT events is given in Supplementary file 3.

1417 **Figure 2. Schematic representation of the structure of HgcAB protein in *P. pyriformis* and its**
1418 **comparison to the prokaryotic HgcA-HgcB operon**. **A)** gives the overview of the full alignment,
1419 which reveals remarkable similarity and conservation in both HgcA and HgcB parts. **B)** Zooms on
1420 the primary structure of the functionally most important regions predicted, namely corrinoid

1421 binding domain (CMD) in HgcA part including the well-conserved corrinoid binding alpha-helix
1422 and cobalamin binding residues (Thr60, Thr66, Val91, Cys93, and Ala153 in *P. mercurii*), an N-
1423 terminal transmembrane domain (with four transmembrane domains in contrast to five in *P.*
1424 *mercurii*), and a conserved motif (NVWCAAGKG) at the positions 97 – 105. The *P. mercurii*
1425 HgcB shares with the *P. pyriformis* sequence the bacterial 4Fe-4S ferredoxin and two strictly
1426 conserved CxxCxxCxxxCP 4Fe-4S binding motifs, conserved C-terminal cysteine residues, and a
1427 cysteine residue (Cys73 in *P. mercurii*) at the beginning of an alpha-helix. **C)** Displays the
1428 complete domain structure.

1429 **Figure 3. Complement of Golgi-associated proteins in Preaxostyla.** This Coulson plot shows
1430 the set of proteins present in the Preaxostyla predicted proteomes. Empty segments denote failure
1431 to identity a candidate orthologue, while filled segments denote success, with paralog numbers
1432 inset as relevant. Candidate proteins are identified by homology-searching and verified by
1433 phylogenetics as relevant. Details are given in Supplementary file 6.

1434 **Figure 4. Preaxostyla_Glycolysis.** Hypothetical reconstruction of the extended glycolysis in
1435 Preaxostyla. Bold outline indicates alternative glycolytic enzymes. Abbreviations and Enzyme
1436 Commission numbers are given in Supplementary file 6. Presence of the enzymes in Preaxostyla
1437 data sets is indicated by a color code.

1438 **Figure 5.** Phylogenetic relationship among hydrogenases of Preaxostyla. A) Schematic
1439 representation of domain architectures of hydrogenases in Preaxostyla species. HydA: [FeFe]
1440 hydrogenase; CysJ: NADPH-dependent sulfite reductase; NuoG: NADH-quinone oxidoreductase.
1441 B) Detailed view of the part of a hydrogenase phylogenetic tree that corresponds to a clade

1442 including long hydrogenases. The domain architecture of the proteins is indicated by color bars.
1443 The full tree is given in Supplementary file 4 U.

1444 **Figure 6.** Summary of the searches for proteins physically associated with MROs. In the upper
1445 panel, the numbers of candidates recovered by the four strategies for the five Preaxostyla species
1446 (color-coded) are summarized. Detailed information on the searches and candidates is given in
1447 Supplementary file 9. In the lower panel, the mitochondrion hallmark proteins detected in the data
1448 sets of *P. pyriformis* and *T. marina* are shown in blue. No such candidate was recovered for any
1449 oxymonad representative.

1450 **SUPPLEMENTARY FIGURE LEGENDS**

1451 **Figure S1.** Hypothetical map of the amino acid metabolism in *P. pyriformis*. Brown color indicates
1452 enzymes possibly involved in amino acid biosynthesis pathways. Red color indicates enzymes
1453 possibly involved in ATP production. Note that some of the connections between metabolites
1454 correspond to the mere transfer of the amino group rather than conversion of the carbon backbone
1455 of the molecule (green color). Cyan color is used for remaining enzymes. Grey outline shows MRO
1456 localization based on Zítek et al. (2022). Abbreviations and Enzyme Commission numbers are
1457 given in Supplementary file 6.

1458 **Figure S2.** Hypothetical map of the amino acid metabolism in *T. marina*. Brown color indicates
1459 enzymes possibly involved in amino acid biosynthesis pathways. Red color indicates enzymes
1460 possibly involved in ATP production. Note that some of the connections between metabolites
1461 correspond to the mere transfer of the amino group rather than conversion of the carbon backbone
1462 of the molecule (green color). Cyan color is used for remaining enzymes. Striped outline shows

1463 putative MRO localization based on comparison with *P. pyriformis*. Abbreviations and Enzyme
1464 Commission numbers are given in Supplementary file 6.

1465 **Figure S3.** Hypothetical map of the amino acid metabolism in *M. exilis*, *B. nauphoetae*, and *S.*
1466 *strix*. Brown color indicates enzymes possibly involved in amino acid biosynthesis pathways. Red
1467 color indicates enzymes possibly involved in ATP production. Note that some of the connections
1468 between metabolites correspond to the mere transfer of the amino group rather than conversion of
1469 the carbon backbone of the molecule (green color). Cyan color is used for remaining enzymes.
1470 Abbreviations and Enzyme Commission numbers are given in Supplementary file 6.

1471 **Figure S4.** Venn diagram showing the distribution of orthologous groups (OGs) of Fe-S cluster-
1472 containing proteins among the species of Preaxostyla. The identity of the OGs and of the
1473 component proteins are provided in Supplementary file 8.

1474 **TABLES**

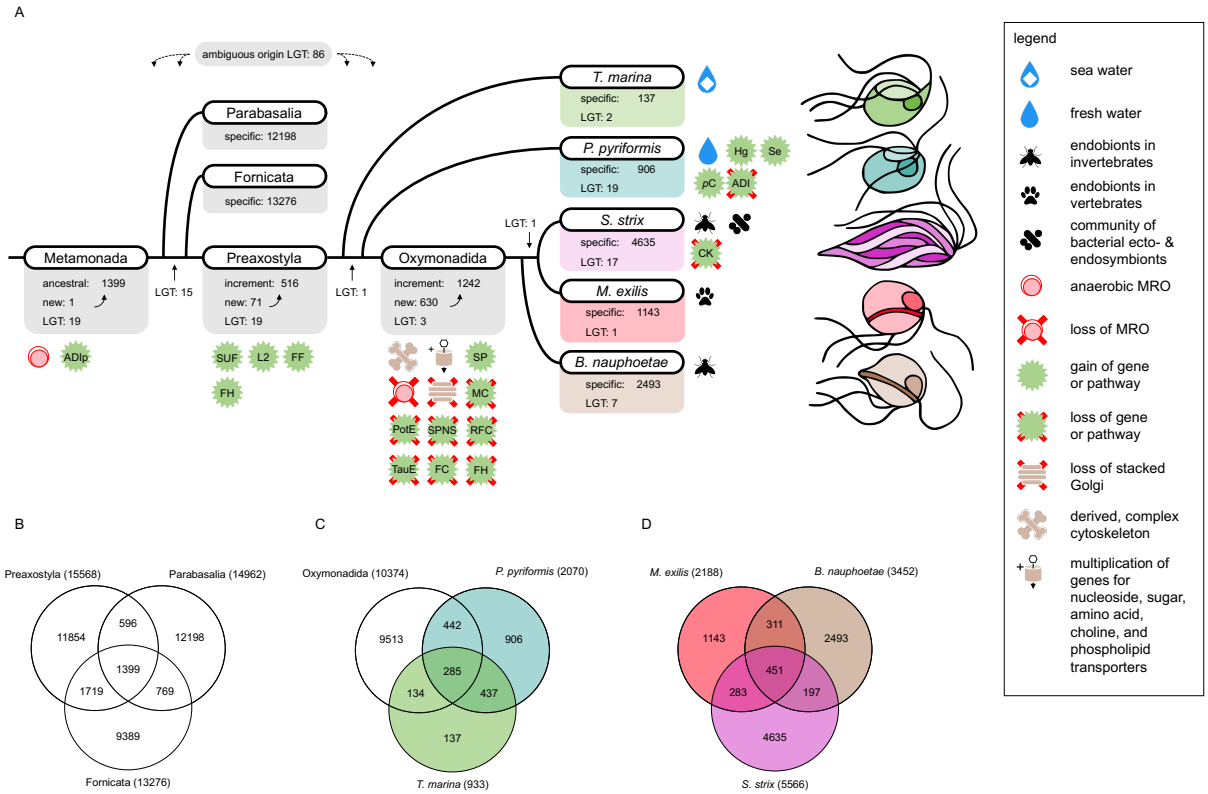
1475 **Table 1. General features of the Preaxostyla genomes discussed in this study**

Sample	Scaffolds	Total length (bp)	N50 (kbp)	Completeness (%; BUSCO odb9)	G+C content (%)	Protein-Coding Loci	NCBI Accession
P. pyriformis	633	56,627,582	276.6	76.3	60.94	10,815	JAPMOS000000000
B. nauphoetae	879	88,537,989	199.5	76.6	44.96	25,221	JARBJD000000000
M. exilis	2,092	74,712,536	71.44	75.3	36.8	16,768	LSRY00000000
S. strix	50,889	152,152,197	5.18	69.6	26.6	56,706	SNRW00000000

1476

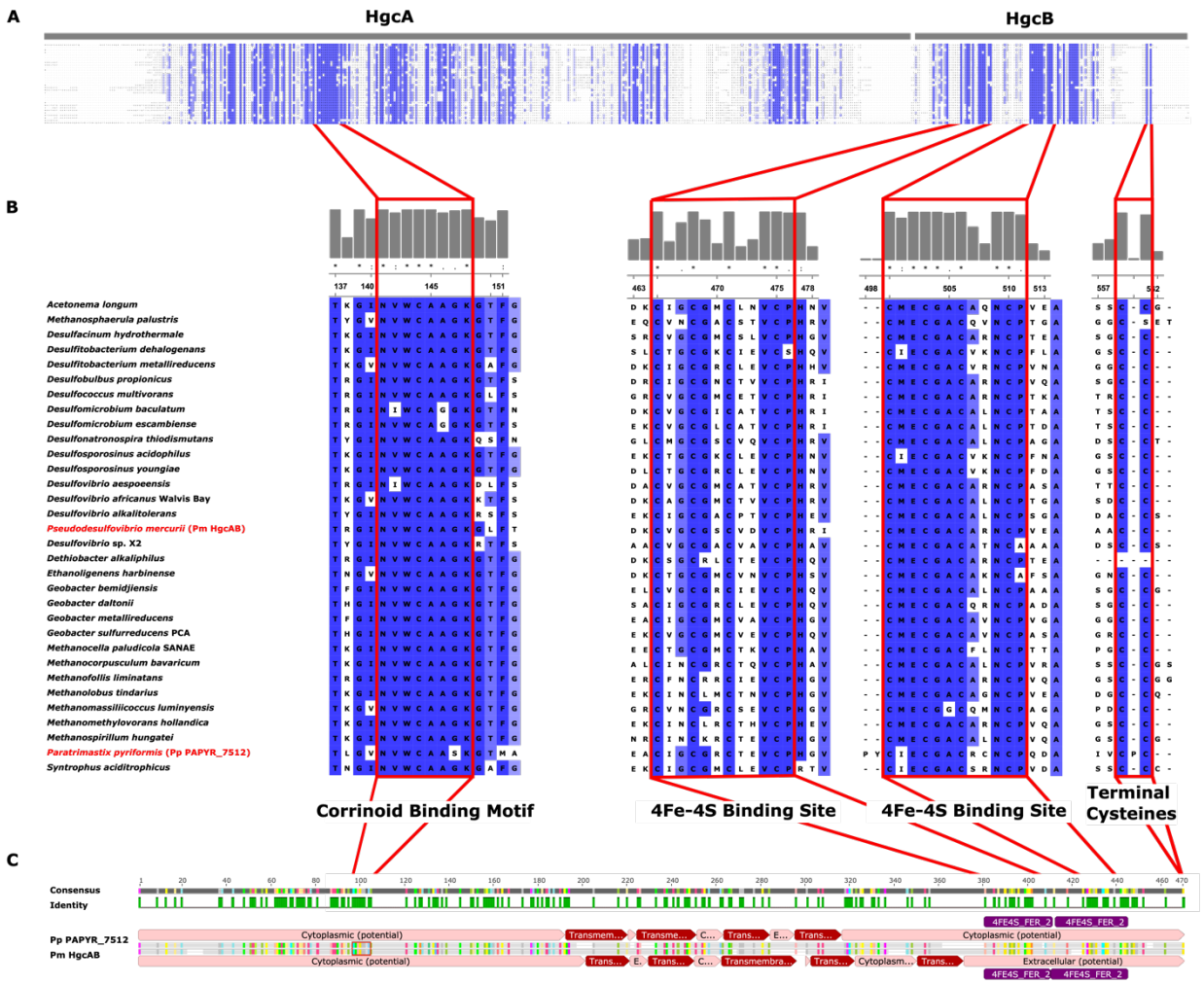
1477 **FIGURES**

1478 **Figure 1**



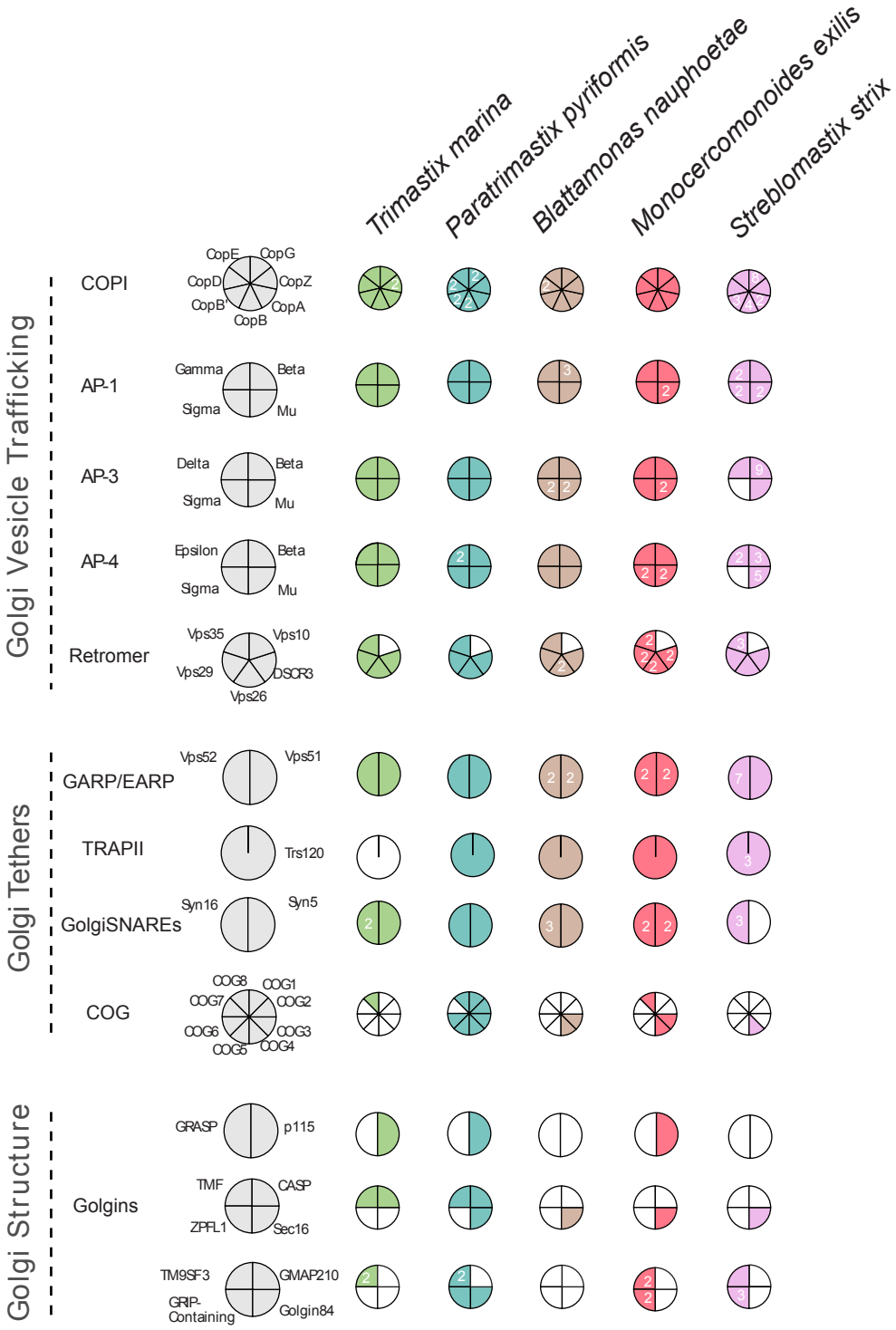
1479

1480 **Figure 2**



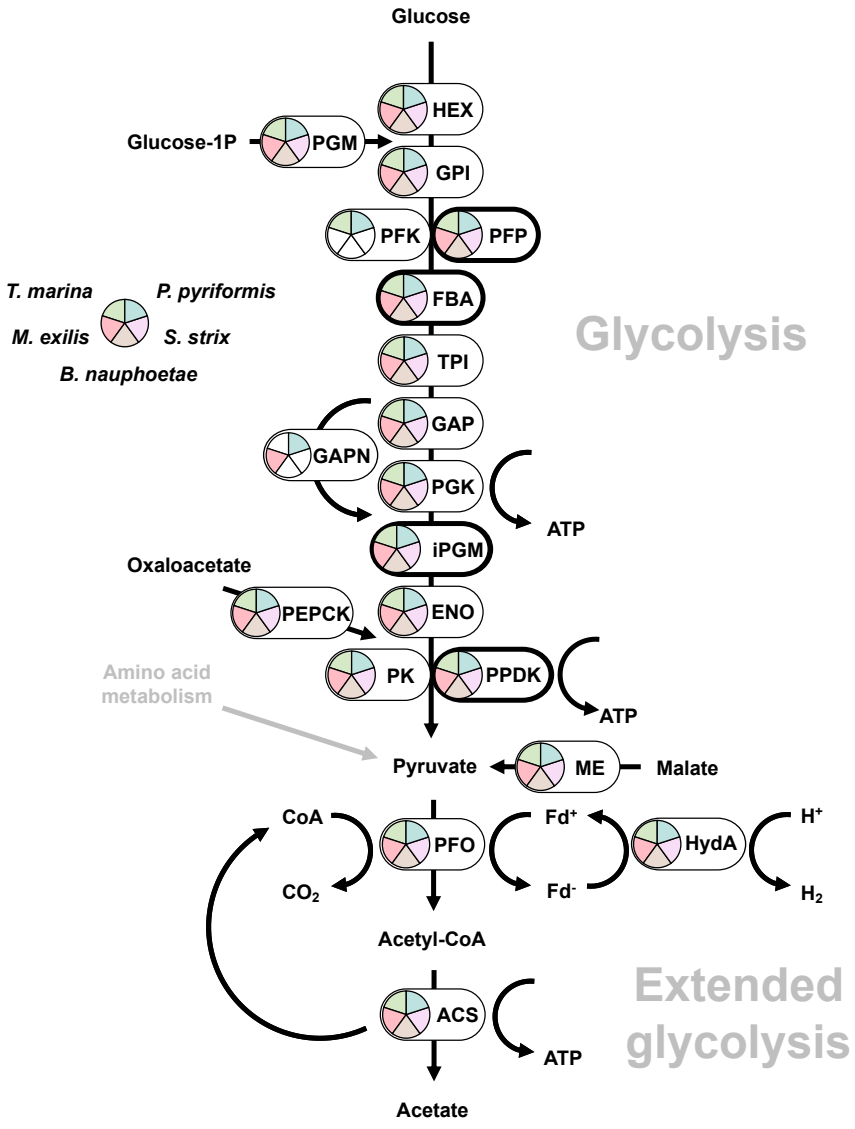
1481

1482 **Figure 3**



1483

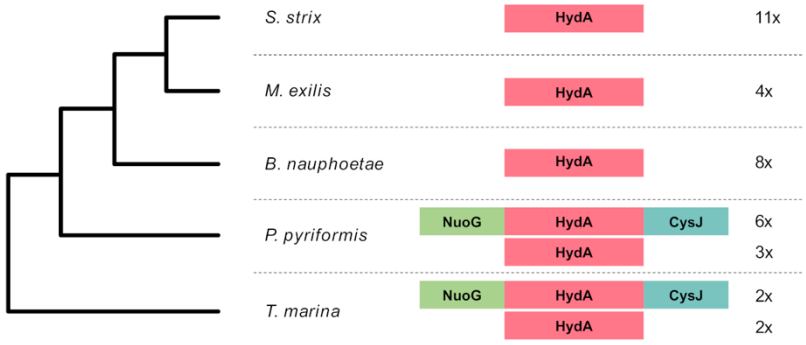
1484 **Figure 4**



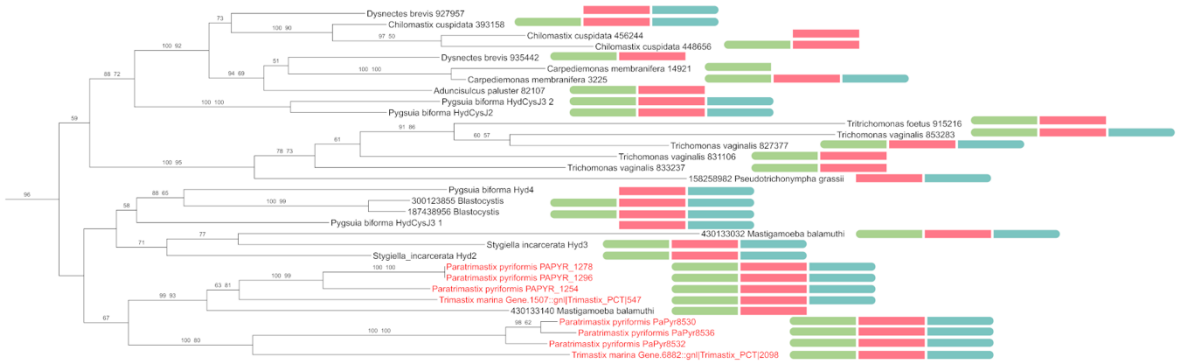
1485

1486 **Figure 5**

A

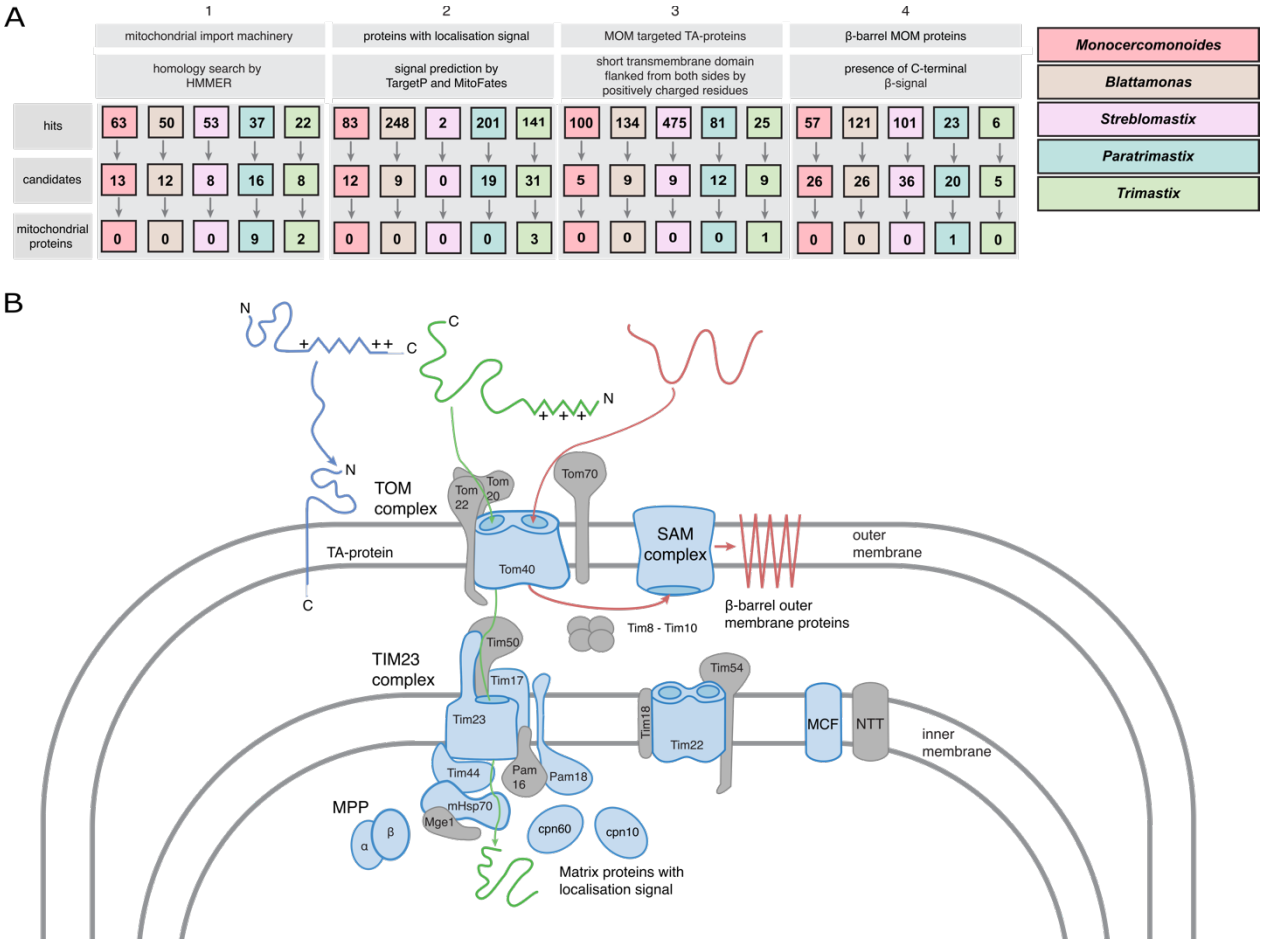


B



1487

1488 **Figure 6**

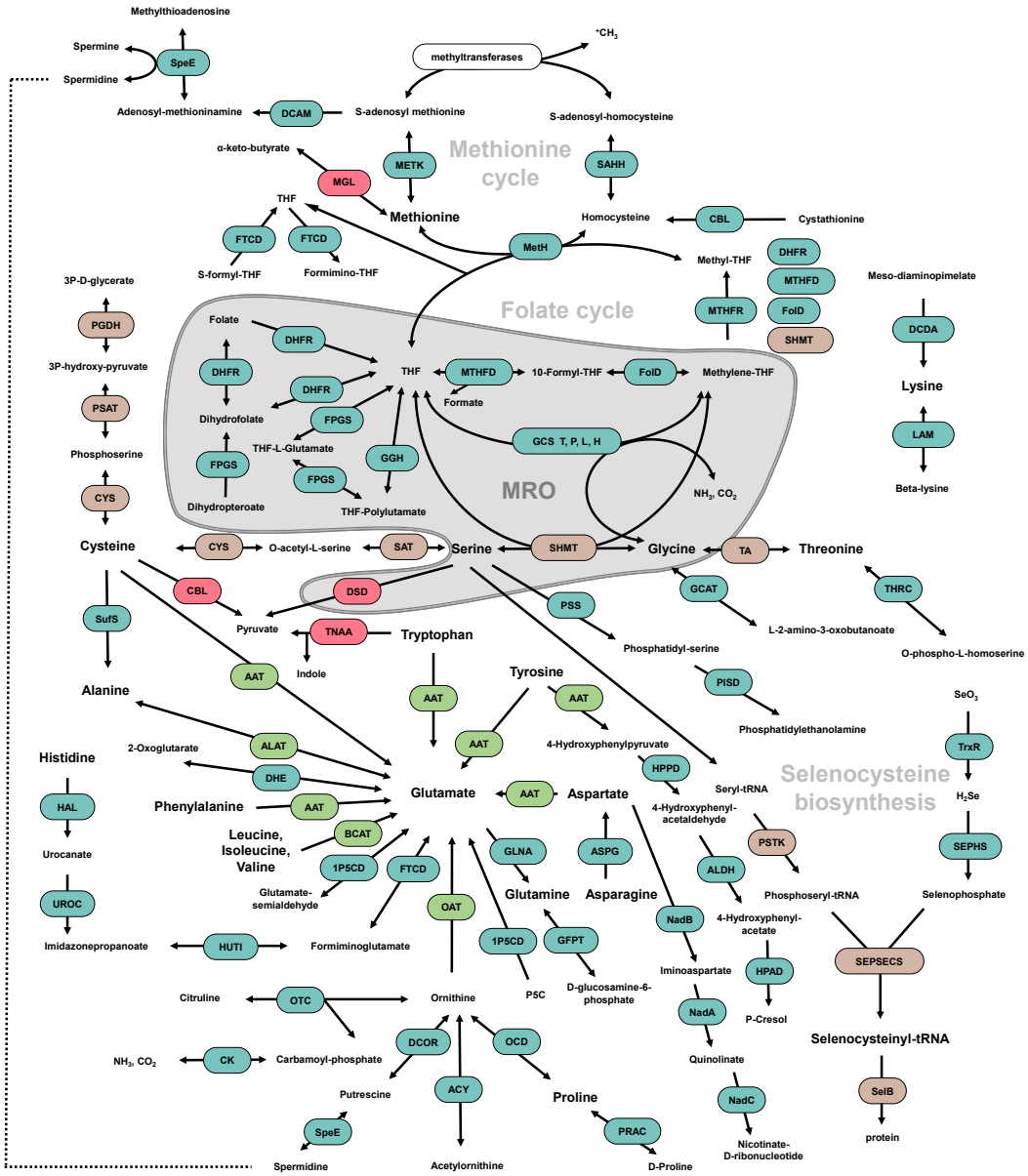


1489

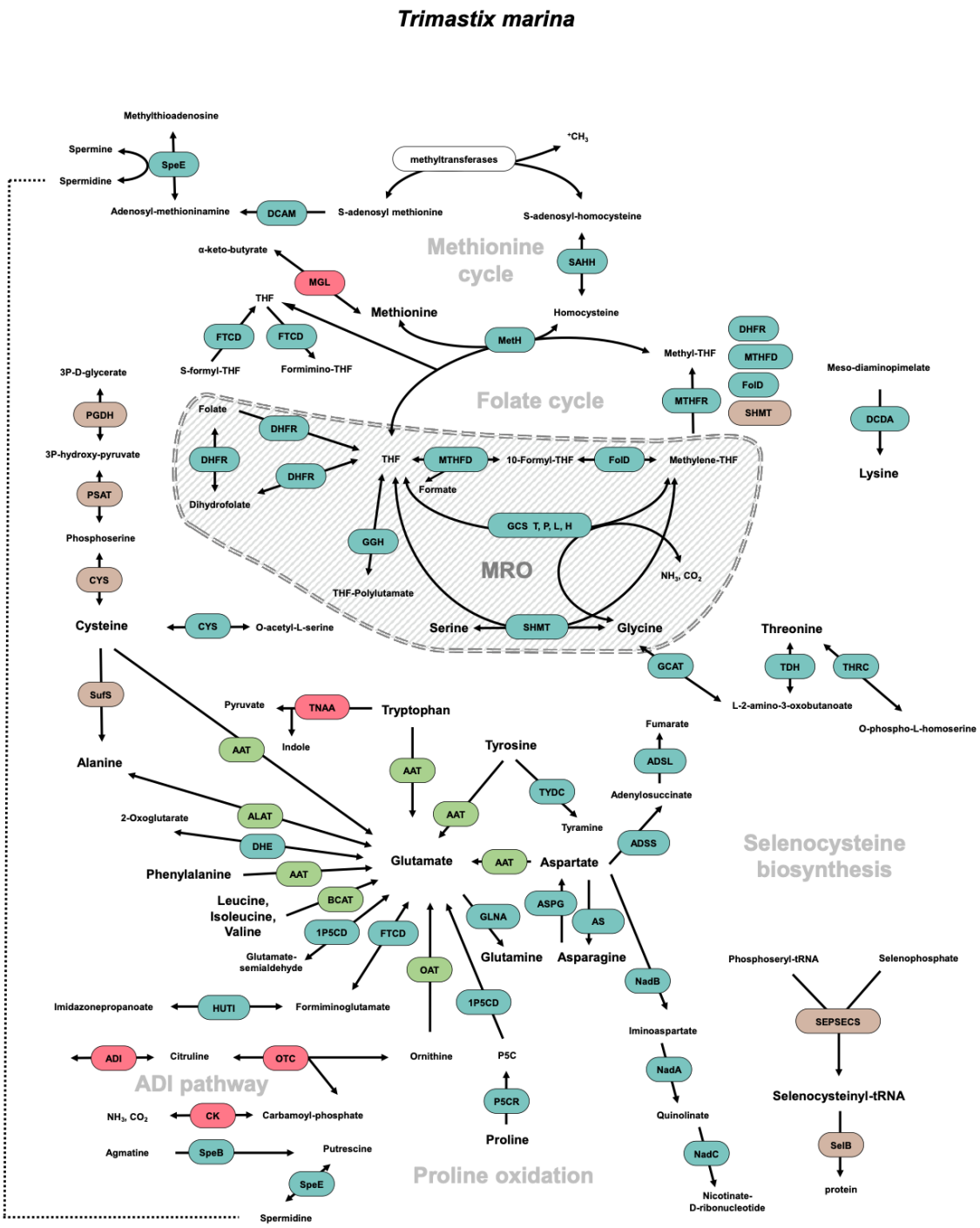
SUPPLEMENTARY FIGURES

Supplementary figure 1

Paratrimastix pyriformis

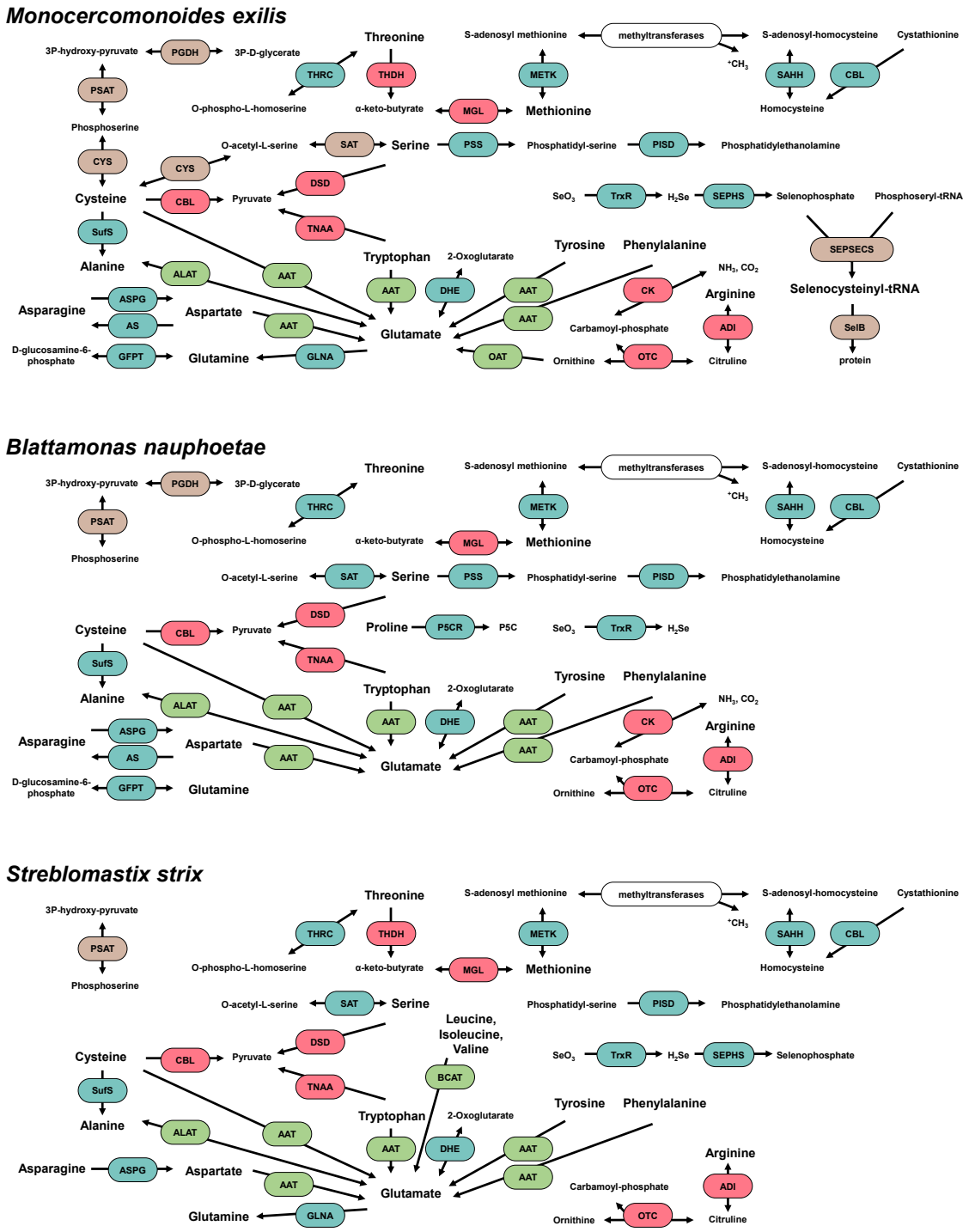


1493 **Supplementary figure 2**



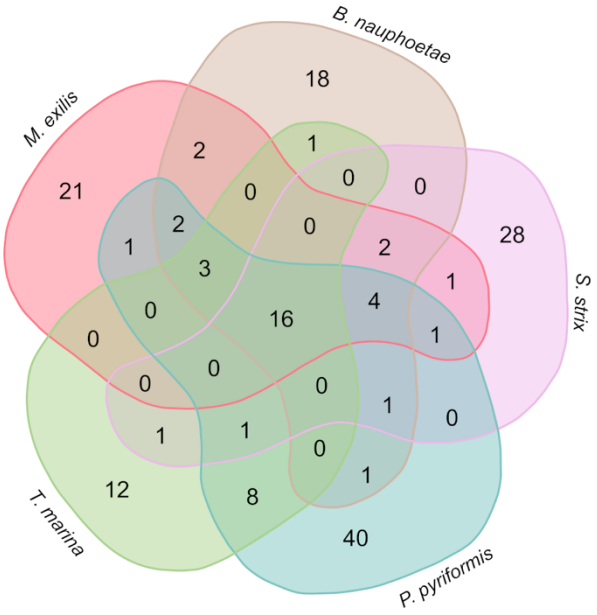
1494

1495 **Supplementary figure 3**



1496

1497 **Supplementary figure 4**



1498